AN INTRODUCTION TO DATA ENVELOPMENT ANALYSIS WITH SOME OF ITS MODELS AND THEIR USES

R. Banker, A. Charnes, W. W. Cooper,
John Swarts, and D. A. Thomas

ABSTRACT

This paper is an introduction to DEA (Data Envelopment Analysis). Basic concepts and different DEA models covered include the "CCR" and "BCC" ratio forms, which are examined in relation to each other as well as in relation to the "Additive" and "Extended Additive" forms that are also included in this discussion. The models presented are accompanied by interpretations which also relate DEA to other disciplines such as economics and management science. Illustrative examples provided in this paper are intended to facilitate understanding and use of DEA in conjunction with, or as alternatives to, presently used methods for evaluating efficiency and controlling the performance of nonprofit and governmental entities.

Research in Governmental and Nonprofit Accounting, Vol. 5, pages 125–163.

ISBN: 0-89232-975-0

I. INTRODUCTION

The three papers that follow this one provide reports on selected uses of Data Envelopment Analysis (DEA) in evaluating the efficiency of organizations engaged in nonprofit and governmental activities as follows: (1) universities, (2) electric cooperatives, and (3) vehicle maintenance units in the U.S. Air Force. The activities in each of these examples involve multiple inputs and multiple outputs with, in general, no "bottom line" available to evaluate performance. Applications in these three papers provide examples of different uses of DEA which include (a) obtaining perspective on findings of a recently completed study of institutions of higher learning in Texas by a legislatively appointed Select Committee on Higher Education (SCOHE), (b) supplying guidance for the management audits of electric cooperatives required by the Texas Public Utility Commission (PUC), and (c) providing a new approach for estimating amounts of "organizational slack" which might be present in vehicle maintenance activities by means of a field experiment conducted with the cooperation of the Tactical Air Command (TAC).

The purpose of the present paper is to provide an introduction to DEA concepts and methods along with some of the models that are now available for implementing DEA studies. Additional help, including validation tests with accompanying interpretations for establishing relevance of DEA results are provided in the next three papers. A still further paper by R. D. Banker is then presented as an introduction to another set of three papers which involve econometric applications within a DEA framework—including hypotheses tests and other statistical analyses. Following this introduction (by Banker) another set of three papers concludes this collection with examples of DEA applications to (1) analyzing cost variances for management control of performance in hospitals, (2) incorporating value judgments for use in efficiency analyses of U.S. Army recruiting activities, and (3) evaluating recent changes of U.S. Air Force accounting and finance offices for their effects on efficiency.

The three papers immediately following our introductory presentation compare DEA with other approaches, such as unit cost or performance ratios, index numbers and regressions used for evaluating efficiency. The next section of this introduction, therefore, tries to supply background and perspective by verbally describing DEA and then comparing some of its properties with what can be expected from these other approaches. Section III then attempts a portrayal of DEA and its workings by a graphical example involving only one input and one output. Starting with Section IV, the models needed for actual use are introduced in what are called "CCR" or "CCR Ratio" forms of DEA with attention directed only to one type of inefficiency in the form of the waste which economists refer to as technical inefficiency. Additional kinds of DEA models are

introduced in subsequent sections to deal with returns-to-scale efficiencies as well as nondiscretionary and threshold or ceiling conditions which may need to be taken into account in arriving at efficiency evaluations.

II. DEA AND ALTERNATIVE APPROACHES

DEA involves an application of mathematical programming to observed data to locate frontiers which can then be used to evaluate the efficiency of each of the organizations responsible for the observed output and input quantities. The solid line shown in Figure 1 of the next section of this paper is a frontier such as might be derived via DEA from data on the amounts of the one input utilized and the one output produced by each of the five entities.

In DEA, the entities responsible for converting inputs into outputs are referred to as Decision Making Units (DMUs). This usage is generic and comprehends the activities of many different kinds of organizations and their subdivisions. Examples provided in the papers that follow include individual universities, electric cooperatives, and Air Force vehicle maintenance units (with the latter operating at subdivision levels of larger entities).

Multiple outputs and multiple inputs may be used in DEA with each being stated in different units of measurement. Cross-comparisons to weight or evaluate the relative importance of these different inputs and outputs are not needed in arriving at evaluations of technical efficiency in DEA because, being wasteful, the related input reductions or output augmentations can be effected without worsening any other input or output. DEA is used (a) to locate the DMUs responsible for these technical inefficiencies, and (b) to identify the sources and amounts of inefficiency in each of its inputs and outputs.[1]

Avoidance of cross-comparisons of different inputs and/or outputs is accomplished in DEA by suitably arranged constraints. Suppose, for instance, that a particular DMU is to be evaluated by reference to the performance of other DMUs using the same inputs and producing the same outputs. For convenience of reference, we designate DMU_o as the DMU to be evaluated. Solution possibilities obtainable from the output and input values observed for all DMUs are then examined with mathematical programming methods and the resulting solutions are used to evaluate the performance of DMU_o. The solutions are required to satisfy constraints which do not allow any of DMU_o's observed input values to be increased or any of its observed output values to be decreased. Mathematically speaking, a properly oriented inequality constraint is imposed to assure that the solutions satisfy this condition for *each* of DMU_o's inputs and outputs.

Inequalities rather than equations are used to allow solutions which can better some inputs or outputs (without worsening other inputs or outputs). To avoid elaborate trial and error or simulation search for such betterment possibilities, the optimization machinery of mathematical programming is used to find a "best" solution. Geometrically interpreted, this best solution will be located on a frontier from which a comparison of DMU_0's behavior can be affected. Because of the constraints, no input or output is worsened. If this best solution does not improve upon any input or output DMU_0 is rated as 100 percent efficient; otherwise it is rated as inefficient and a straightforward reading of the solution locates the sources and amounts of inefficiency in each input and output.

Only *relative* efficiency evaluations are obtainable when solutions are generated from observed data in this manner.[2] The solutions generated in this manner do not generally result in input and output values that are coincident with the observed behavior of any actual DMU. Nevertheless, differences between the solution and DMU_0's observed input and output values are said to provide "evidence"[3] of its inefficiencies because these solutions are obtained from observed data.

At this point, it is useful to compare what has just been said with what is done in other approaches that are now in use. Ratios in the form of unit costs provide one example. The use of a cost-per-student ratio as one criterion for evaluating university performance in Table 2 of the immediately following paper is an example in which each university is judged relative to an average of such unit costs. The value of this average which is used as a criterion of performance need not conform to the actual unit cost of any university.

Extensions to deal with multiple outputs and inputs, and continuation of a ratio analysis approach can result in arrays like the one shown for the San Patricio Electric in Table 1 in the second of the two articles succeeding this one. Prepared by the Rural Electrification Administration (REA), a total of 670 ratios and other averages is made available in this manner for possible use in evaluating San Patricio's performance. See the discussion of this Table that is given in the article and which, in turn, is based on a more detailed development in Thomas (1986).

To avoid or to complement such complex arrays, recourse might be made to synthesizing an "index number" as a single summary number—e.g., for use by the Texas Public Utility Commission in identifying candidates for "efficiency audits" from among the 75 electric cooperatives for which such audits have been legislatively mandated. Synthesis of such an index number would, however, require recourse to a priori weights to reflect the relative importance of its component costs and volumes in order to obtain a suitable efficiency measure. Such an effort could encounter numerous difficulties that are bypassed by DEA because the latter does not require such a priori weight selections in order to arrive at its overall efficiency values. Instead, the "weights" to be assigned to DMU_0 are obtained automatically from the data as part of the solution to the mathematical programming problem used to effect DEA efficiency evaluations.

The DEA literature uses the term "virtual transformations" for these parts of the mathematical programming solution in order to avoid confusion with customary weightings. Mathematically speaking, these virtual values transform DMU_0's observed outputs and inputs into a "virtual output" and a "virtual input" which in the form of a ratio of virtual output-to-virtual input provides a measure of DMU_0's efficiency for use with the CCR or BCC ratio forms of DEA. This is all accomplished by the computer codes used to apply the models to the observations (for all DMUs). The term "virtual" distinguishes these derived values from actual observations and the resulting ratio of virtual output-to-virtual input may be regarded as an extension of the usual output to input form used for productivity (or efficiency) indexes. Unlike the usual index number usage of fixed weights, however, the values assigned to these virtual transformations by the mathematical programming solutions of DEA depend on the mixes of outputs and inputs used by each DMU_0 and, indeed, the values obtained in this manner are optimal in the sense of giving the best possible virtual output-to-virtual-input value for each DMU_0 that is evaluated.

Index numbers with fixed weights are not the only alternative. A use of statistical regressions represents another alternative and this approach can also allow for differences in mixes and volumes in arriving at efficiency evaluations that are applicable to different DMUs. Exhibit 1 in the second of the following three papers provides an example in the form of nine separate regressions calculated annually by REA and supplied to individual electric cooperatives for use in evaluating their performance in each of the activities covered by these regressions. Calculated from data in a national sample of electric cooperatives, the resulting regressions are assumed to be "representative" so that estimates obtained from these regressions can be used to provide unbiased estimates of performance in the following manner. A particular cooperative like San Patricio can insert its particular values for each of the independent variables listed in the stub of Exhibit 1, and use the regression relation to obtain an estimate of the value of the corresponding dependent variable. Actually, this is all done by REA and compiled in an arrangement like the one shown in Exhibit 2 for use by a cooperative (as well as the REA field office) in evaluating its performance. Help which is needed to interpret these results is supplied in a variety of forms including the heavy black dots, which are used in Exhibit 2 to flag items that appear to be out of line.

Even though these dots are oriented for attention to efficiency, the regressions used do not make any separations between efficient and inefficient performances that may be present in the data base. To state this differently, the regression estimates are formed from observations which contain inefficient and efficient behavior in unknown proportions. This is in contrast to DEA where a clear separation of efficient and inefficient behavior is effected because the mathematical programming solutions use only frontier estimates to effect efficiency evaluations.

Statistical estimates need not be derived only in the form of "central tendencies," as in the usual least-squares regression approaches. They can also be oriented toward frontiers by suitably oriented constraints as in the work by

Aigner and Chu (1968). In this form, they can also be used to validate any use of central tendency approaches as was done by Charnes, Cooper and Sueyoshi (1988), for example, in their review of a study commissioned by the U.S. Justice Department for use in its antitrust action against AT&T.[4]

Whether used to secure central tendency or frontier estimates, however, these regression approaches are heavily dependent on numerous assumptions that must be made including assumptions about causal relations between outputs and inputs that need to be specified explicitly.[5] A change from the assumption that these relations are linear to an assumption of a Cobb-Douglas (or other log linear) form, for example, can produce very different results from the same data. See Dewald, Thursby and Anderson (1986) and also Lalonde (1986) on the need for checking regression results against alternate possibilities.

As noted in Charnes, Cooper, Golany, Seiford and Stutz (1985), DEA does not require explicit specification of the functions that are supposed to relate outputs to inputs. There are theoretical reasons for believing that results in the form of characterizations of efficient and inefficient behavior will be relatively robust even with DEA model changes—e.g., from CCR to BCC or other forms (see Ahn, Charnes and Cooper, 1989; Charnes and Zlobec, 1989). However, a variety of different DEA models are included in later sections of this paper both for perspectives on DEA, and because it may be desirable to check results secured from one DEA model against other possibilities.

We also need to emphasize that the orientations in DEA are toward the observations associated with individual DMUs. Thus, in contrast with customary uses of statistical estimates, which are oriented toward *all* observations, DEA introduces a new principle for effecting estimates from empirical data which is oriented toward *each* observation. A least squares regression of the usual statistical variety, for example, uses only a single optimization to obtain a single estimating relation from n observations, whereas DEA uses n optimizations for the same n observations in order to obtain efficiency evaluations for each DMU. Stated differently, the thus derived (single) regression is assumed to be applicable to each DMU, as in the examples of the Table 2 use of the REA computed regression to evaluate the performance of the San Patricio electric cooperative. No such assumption is made in DEA and indeed the underlying functions may differ from one DMU to another.

Because of the absence of any need for prescribing the underlying functional forms, weights, etc., in an a priori manner, it is fair to say that DEA is "empirically based" in contrast to uses of statistical regressions, productivity indexes, etc., which, in principle, require a great deal of analytical theorizing prior to choosing the forms in which they are to be used. This empirically based orientation has provided an opening for addressing many problems which have not heretofore received much research attention and it has also provided new ways of addressing problems which have been researched by other methods. As noted in

the concluding portion of the present paper, this orientation in DEA makes it easier to attend to issues of relevance (as well as validity) of results by using study strategies which invite the participation of managers in choices of the DMUs as well as the inputs and outputs to be used.

III. GEOMETRIC PORTRAYAL

Figure 1 will help to portray what is involved in a use of DEA by means of a single output-single input example. For this geometric depiction, we associate the points P_1, P_2, P_3, P_4, P_5 with Decision Making Units (DMUs) designated as DMU_1, DMU_2, DMU_3, DMU_4, DMU_5. Using y_j for the output and x_j for the input associated with DMU_j, the values of these outputs and inputs can be regarded as coordinates of points P_j in order to represent geometrically the observed behavior of the corresponding DMU_j, $j = 1, \ldots, 5$. (See each of the y_j and x_j pairs which are parenthesized as coordinates alongside the point P_j to which they refer in Figure 1.)

How can performance efficiencies be determined from these observations? One possibility is to effect comparisons by reference to an output per unit input ratio for each DMU. For instance, DMU_1 which is associated with the point P_1

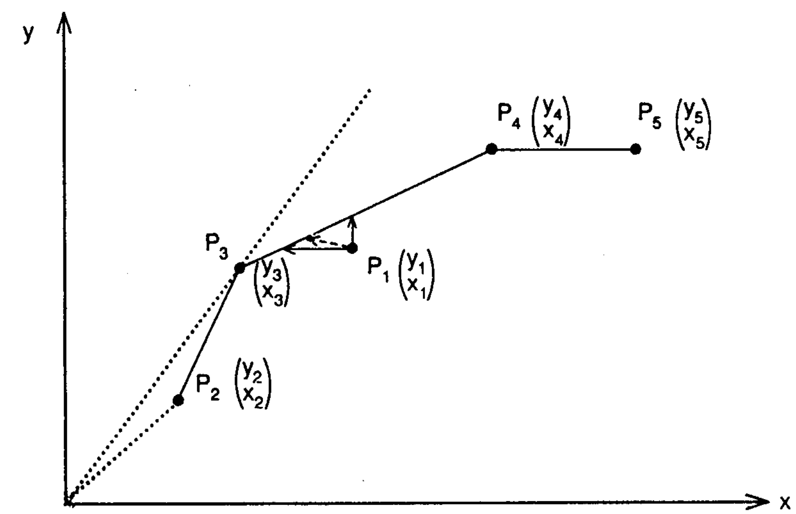

Figure 1. Efficiency Evaluations in a One Output–One Input Example

might be evaluated in an ordinary ratio approach by comparing the ratio y_1/x_1 with an average value for all 5 ratios calculated via

$$\frac{\sum_{j=1}^{n} y_j/x_j}{5} = \frac{y_1/x_1 + y_2/x_2 + y_3/x_3 + y_4/x_4 + y_5/x_5}{5}$$

One could then determine whether y_1/x_1 is greater or less than the thus computed average of these ratio values as a way to assess the relative efficiency of DMU_1 in producing its observed output y_1 from its observed input x_1. A problem can arise as to whether inefficient points like P_1 should be allowed to enter into an average that is to be used to determine whether the performance of P_1 is efficient and, of course, there will generally be more than one such inefficient point to be considered. Even if distinctions are to be made in point by point fashion there is no criterion that is evidently available to effect the distinctions between efficient and inefficient performers. Notice, for instance, that $y_2/x_2 = y_1/x_1$ even though P_2 is associated with an efficient DMU and P_1 is not.

Another problem can arise because the output and input magnitudes, as well as their ratios, may require consideration, and this will generally lead to a need for some kind of a priori system of weights. Finally, when it is necessary to deal with multiple outputs and multiple inputs, these weighting schemes will need to be extended with accompanying complications.

An alternate possibility is to confine comparisons to DMUs which are "closest to" (or "most like") DMU_1 in their output and input mixes. DEA carries this a step further by effecting its evaluations by reference to subsets of efficient DMUs so that, *inter alia*, data from inefficient performers are excluded from the comparison set.

To explain how this is accomplished by DEA, we begin by noting that a necessary condition for a DMU to be efficient (for DEA) is that the point representing its observed output and input values must be on a frontier like the one indicated by the solid line connecting the points P_2, P_3, P_4 and P_5 in Figure 1. In fact, P_3 and P_4 represent the subset of observed efficient points which are closest to P_1. The latter (i.e., P_1) is not on the frontier and, hence, DMU_1 can be judged inefficient relative to the frontier as can be seen in Figure 1.

A problem arises in using P_3 and P_4 for evaluating P_1 if it is desired to avoid cross comparisons with accompanying assumptions as to the relative values of inputs and outputs. Note that even though P_3 and P_4 are efficient, their output and input values are not both "better" than those for P_1. P_4 has an output which is "better" than that recorded for P_1, in that $y_4 > y_1$, but it also has an input value which is "worse," in that $x_4 > x_1$. Similarly, for P_3 we have $x_3 < x_1$ but also $y_3 < y_1$. DEA resolves this problem by allowing all points on the line connecting P_3 and P_4 to form part of an "efficiency frontier." These points are

then considered to be available for use in obtaining the desired efficiency comparisons for P_1 even when the coordinates of the point from which the comparisons are to be made do not represent actually observed input and output values.

Location of the line segment connecting P_3 and P_4 as part of the pertinent "facet" of the efficient frontier for use in evaluating P_1 still leaves open the choice of a particular comparison point for determining the sources and amounts of inefficiency reported for DMU_1's behavior. Any point between the solid vertical and horizontal arrows leading from P_1 to this efficient facet yields a point with coordinates which are both at least as good as the values for DMU_1. That is, any such point will have coordinates y and x which satisfy $y \geq y_1$, and $x \leq x_1$ so that the output and input values for such a point are *both* at least as good as the P_1 coordinates.

One possible choice for comparison is indicated by the point at the end of the vertical straight line and another is indicated by the horizontal straight line stretching from P_1 to the efficiency frontier. The vertical line yields a point on the efficiency frontier for which $y > y_1$ and $x = x_1$ so that all of the inefficiency appears as a difference in the output obtainable from the input amount that was utilized. That is, this choice assigns all of the inefficiency in P_1 to the output value. In similar fashion, the horizontal line assigns all of the inefficiency to the input value. These, however, are not the only choices available for evaluating P_1. Another possible choice is provided by the arrow indicated by the broken line stretching from P_1 to the frontier. This choice locates a point with $y > y_1$ and $x < x_1$ and these output and input differences $(y - y_1)$ and $(x_1 - x)$, are interpreted as amounts of both output and input inefficiencies in the observed performance of DMU_1.

Different DEA models may be used to obtain different comparison points but, as shown in Ahn, Charnes and Cooper (forthcoming, 1989a), the characterizations of full efficiency (i.e., 100% efficiency) will generally be the same with different models.[6] Moreover, as shown in Charnes and Zlobec (forthcoming, 1989), efficiency characterizations in DEA are also generally stable in the presence of perturbations in the frontiers. By recourse to these analytical-theoretical developments, it is therefore possible to provide assurance that DEA rests on a body of underlying concepts and methods which can be used with different model forms, at least as far as efficiency characterizations are concerned, without encountering the often drastic differences that accompany different choices of ratios and/or regressions (see Ahn, 1987, pp. 45–47).

Inefficiencies like those discussed for P_1 are associated with what are called "technical inefficiencies." Such inefficiencies, when present, are identified by DEA in the manner that has just been indicated—namely, the evidence shows that an output augmentation or an input reduction can be effected without worsening any other output or input. Another kind of inefficiency is so-called scale (or returns to scale) inefficiency and this kind of inefficiency can also be made part of a DEA analysis. Consider, for instance, the broken line ray from the origin to P_2 in Figure 1. Movement on the frontier from P_2 to P_3 will be associated with increases in the slope of the corresponding rays. This means that average output is increasing with increasing amounts of input; hence, this evidence indicates that increasing returns to scale are occurring on this facet of the frontier.

These increases continue until P_3 is reached. Movements from P_3 to P_4, on the other hand, are associated with decreases in the slope of the ray from the origin to the frontier so that decreasing returns to scale are occurring with each increase of input on this segment (or facet) of the frontier. These returns-to-scale characterizations, it should be noted, refer only to movement on efficient frontiers. The concept of returns to scale is ambiguous when applied to points like P_1 where technical inefficiencies are also present. Indeed, the definition of a production function as used in economic theory is oriented toward frontiers. This means that, given a production function, symbolized as f(x), the output y obtained from any specified x must be maximal. Hence, if x is chosen equal to x_1 and y_1 is the observed output, we must have

$$\frac{y_1}{y^*_1} \leq 1 \tag{1}$$

where y^*_1 is the theoretically maximal output obtained from this choice of $x = x_1$. See Rhodes (1978) for more detailed discussions and see Section 4, below, for how returns to scale are determined empirically with the BCC and CCR models of DEA.[7]

Evidently (1) can be used to determine technical inefficiency. For, when knowledge of the production function is available, the value of $y_1{}^*$ can be determined from x_1 and the difference between $y_1{}^*$ and y_1 would represent the amount of technical inefficiency observed for DMU_1. Generally speaking, however, the knowledge which economic theory requires for use of a true production function is not available. Nevertheless, concepts associated with production function theory in microeconomics can still be used for guidance. For example, improvements in input consumption or output production are of interest in microeconomics only when they have some positive value or price. It is then assumed that excess inputs will not be used.

It is the assumption that each input and each output has "some" value that makes the location of technical inefficiencies of interest in DEA. In DEA, however, it is not *assumed* that such inefficiencies will *not* occur and, indeed, the location of such inefficiencies is a major objective in most DEA analysis. As noted when discussing P_1, the DEA evaluations for any DMU_o are conducted relative to frontiers. Achievement of a position on the frontier, however, is not sufficient to eliminate technical inefficiency unless the point representing the outputs and inputs of DMU_o is also on an efficient portion of the frontier. Our

simple one input-one output example can help to indicate what is being said if we return to Figure 1 and consider P_5, the point which records the output y_5 and the input x_5 for DMU_5. Evidently, $y_4 = y_5$ so that the output values for P_4 and P_5 are the same but $x_4 < x_5$ so DMU_5 is not efficient. If input has any positive value (or price), no matter how small, the evidence in this example would indicate that DMU_5 is not efficient. It should have been able to reduce its input from x_5 to x_4 without reducing its output and, hence, a wastage of this resource was recorded in the amount of the difference between x_5 and x_4.

The evidence generated by the collection of DMUs represented in Figure 1 shows that both DMU_1 and DMU_5 were "technically inefficient." To extend this to the case of multiple outputs and multiple inputs, we can phrase what is involved as follows: Technical inefficiency is present in the observed behavior of some DMU if and only if the evidence indicates that any of its inputs or any of its outputs an be improved *without* worsening any other input or output. This accords with the already noted assumption that an improvement in any input or output is desirable because all inputs and all outputs are assumed to have "some" positive value. See the discussion of expression (2) in the next section where it is shown how to represent the fact of having "some" positive value without any need to specify this value numerically.

Technical inefficiency, when present, may be said to indicate the availability of a "free lunch." See Leibenstein (1976) and Stigler (1976) for a discussion of the relative importance of such inefficiencies in economics. This kind of inefficiency is thus to be distinguished from scale inefficiency in that achievement of the latter generally requires some input alteration in order to obtain an output increase (or decrease) that will eliminate the scale inefficiency. Thus, movement along an efficiency frontier to achieve scale efficiency generally carries with it a relative imputation of output and input values in the form of prices (or other weights) in order to determine whether (and by how much) such scale return possibilities should be exploited.

A need for recourse to prices or weights to obtain efficiency evaluations creates a difficulty for governmental and nonprofit DMUs because prices or weights are generally not available for all inputs and outputs. This difficulty is avoided by DEA in the manner already noted when considering technical inefficiencies. It needs to be considered anew, however, in identifying scale inefficiencies, especially in multiple output situations because some output-input combinations may be exhibiting increasing returns while others are exhibiting decreasing returns in the same DMU.

One possibility is to identify a point like P_3 in Figure 1 where it can at least be said that increasing returns cease (in going from P_2 to P_3) and decreasing returns begin (in going from P_3 to P_4). This, however, applies to the case of a single output, and such single output-single input cases will not always be available. Nevertheless, a point like P_3 may be identified with constant returns to scale in a manner that extends to the case of MPSS (Most Productive Scale Size) which is a concept introduced in DEA by Banker (1984) to deal with the case of multiple outputs (see also Banker, 1980b).

There are, of course, a variety of complicating issues to consider. Nevertheless, the concept underlying Most Productive Scale Size can be brought into view via the simple example of Figure 1 where, as can be seen, $y_3/x_3 \geq y_j/x_j$ for all j. Hence, the returns to scale associated with the ray from the origin to P_3 is at least as great as at any point. Banker's MPSS concept extends this to the multiple output-multiple input case by proportionately increasing all inputs to a point where decreasing returns begin to appear in at least one output. Carrying an activity beyond this point implies some use of relative prices or weights in order to compare the possibly greater gains in some outputs against the failure to achieve these gains in other outputs.

In some DEA models, such as the CCR ratio form (which is described in the next section) efficiency evaluations may be effected by reference to points on the ray of maximal slope. This means that technical and scale inefficiencies are being considered simultaneously. When desired, however, the analysis may be modified, as in Banker's (1984) article on MPSS or an extension may be made by introducing new variables to separate these different efficiencies as in the BCC models of Banker, Charnes and Cooper (1984) that are discussed in Section IV.

Evidently different DEA models can be used for different purposes. All of these different DEA models are oriented toward frontier concepts, however, and effect their efficiency evaluations for any DMU_o by choosing subsets of efficient DMUs for this purpose. These different models also utilize multiple optimizations, one for each DMU, and their estimating principles therefore differ from regression (and like approaches), and thus the inferences made from the results secured need to take this into account. For instance, suppose a regression were fitted to the points of Figure 1 in the usual least-squares manner and produced a result which indicated that increasing (or decreasing) returns were present. Logically speaking, the presence of statistically determined properties such as increasing or decreasing returns should be considered as only a *class* property so that the application of this result to any *individual* DMU is not justified without further analysis. This differs from what is logically justified from a DEA analysis in that results from the latter analysis are intended to be applicable to *each* DMU_o that is evaluated.

IV. CCR RATIO MODELS AND TECHNICAL INEFFICIENCIES

For purposes of computation and implementation it is necessary to turn to mathematical formulations. We start with the following pair of problems which are a

dual pair of linear programming problems that can be used to obtain efficiency estimates for the CCR ratio form of DEA,

$$\min h_o = \theta_o - \epsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \qquad \max y_o = \sum_{r=1}^{s} \mu_r y_{ro} \tag{2}$$

subject to

$$0 = \theta_o x_{io} - \sum_{j=1}^{n} x_{ij}\lambda_j - s_i^-$$

$$y_{ro} = \sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+$$

$$0 \leq \lambda_j, s_i^- s_r^+$$

subject to

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

$$\sum_{r=1}^{s} \mu_r y_r - \sum_{i=1}^{m} \nu_i x_{ij} \leq 0$$

$$\mu_r \geq \epsilon$$

$$\nu_i \geq \epsilon$$

$s_i^-, s_r^+, \lambda_j, i = 1, \ldots, m; r = 1, \ldots, s; j = 1, \ldots, n.$

As was true in Figure 1, data are assumed to be available for each of j = 1, . . . , n DMUs for use in either of the above two models. Here, however, allowance needs to be made for the presence of multiple inputs and multiple outputs. This is done by introducing an index i for the i^{th} input and an index r for the r^{th} output. The observed input and output values for each DMU_j are then incorporated in the above constraints as follows:

$$x_{ij} = \text{amount of input i for } DMU_j$$

$$y_{rj} = \text{amount of output r for } DMU_j \tag{3}$$

All inputs and outputs are positive.[8] Since these x_{ij} and y_{rj} values are observations, this amounts to assuming that each DMU was observed to use the same i = 1, . . . , m inputs and produce the same r = 1, . . . , s outputs, in possibly different amounts—with the amounts used and produced being indicated by the values of x and y.

The symbol ϵ represents a positive constant introduced as an artifact to insure that all of the observed inputs and outputs will have "some" positive value assigned to them. This value, which need not be prescribed explicitly, serves as a lower limit for the values that can be assigned to the variables μ_r and ν_i as shown by the final constraints in the problem on the right. The variables μ_r and ν_i are the virtual transformations referred to in Section II, above, and the fact that ϵ is positive but not otherwise specified means that each of ν_i and μ_r will have "some" positive value. Mathematically speaking, the value of ϵ is defined to be so small that no multiple of ϵ, however large, can compensate for a reduction in the value of min. $\theta_o = \theta_o{}^*$ where $\theta_o{}^*$ is the optimal value in the left-hand problem. In other words, these $\epsilon > 0$ choices are defined so that the optimal value of θ_o will not be affected by any value that may be assigned to the slack variables associated with ϵ in the objective of the problem on the left.[9]

The s_i^- and s_r^+ are slack variables and the y_{ro} and x_{io} values shown on the left in (2) represent observed output and input values for DMU_o, the DMU being evaluated. That is, the y_{ro} and x_{io} like the y_{ij} and x_{ij} are all known constants. The values that can be assigned to the slack variables in any solution are constrained to be non-negative. Hence, every solution must satisfy $y_{ro} \leq \sum_{j=1}^{n} y_{rj} \lambda_j$ so that efficiency comparisons and evaluations will be effected only from solutions with output values at least as great as the outputs achieved by DMU_o in every case. Similarly, $s_i^- \geq 0$ means the solutions will satisfy $\theta_o x_{io} \geq \sum_{j=1}^{n} x_{ij} \lambda_j$ for each of the i = 1, ..., m inputs utilized by DMU_o. As shown in Charnes, Cooper and Rhodes (1978), an optimal $\theta_o = \theta_o^*$ will never exceed unity so it follows that $x_{io} \geq \theta_o^* x_{io} \geq \sum_{j=1}^{n} x_{ij} \lambda_j^*$. Hence, every observed input amount x_{io} utilized by DMU_o will be at least as great as the one used in its evaluation via $\sum_{j=1}^{m} x_{ij} \lambda_j^*$. Thus, as explained in the discussion of Figure 1, the DEA efficiency evaluations of DMU_o are to be conducted by reference to solutions that do not exhibit reductions in any of its outputs or increases in any of its inputs. Stated differently, the constraints do not allow solutions which involve output or input exchanges when the focus is on technical efficiency.

Efficiency evaluations associated with these solutions will be dependent on the number of degrees of freedom that are available. There are m + s constraints to be satisfied in the problem on the left in (2) and n observations, one for each of the j = 1, ..., n DMUs that form the possible combinations from which efficiency evaluations can be secured. From degrees of freedom considerations, the number of variables λ_j used for the solutions in the problem on the left should be at least as great as the number of constraints. Thus, the number of DMUs for which there are observations should be greater than the number of constraints and, for DEA efficiency evaluations, it is generally advisable to have $n \geq 3(m + s)$. This is only a rule of thumb, of course, which may need to be adjusted in particular situations.

We now formalize the concepts of "virtual output" and "virtual input" as introduced in the preceding section. This will enable us to interpret the solutions to these mathematical programming problems in a slightly different manner. Any DMU from the set of j = 1, ..., n may be singled out for evaluation as a DMU_o. In the problem on the right in (2) its evaluation is accomplished as follows. DMU_o's outputs y_{ro}, r = 1, ..., s, are positioned in the functional at the top to define a new "virtual output," $y_o = \sum_{r=1}^{s} \mu_r y_{ro}$, by selecting virtual weights μ_r which maximize the value of this virtual output, subject to the unity condition imposed on the "virtual input" defined via $x_o = \sum_{i=1}^{m} \nu_i x_{io} = 1$.

Drawing this all together, we interpret the problem on the right as follows: Select values for the virtual transformations, μ_r and ν_i, which will maximize DMU_o's virtual output subject to the condition that its virtual input is equal to unity. The values of the virtual transformations should take the observations of all DMUs into account in a manner that does not permit their transformed outputs to exceed their transformed inputs. This condition is also applicable to DMU_o so that $y_o \leq x_o = 1$. Using stars to denote optimal values, DMU_o is efficient relative to the observation set if an only if $y_o{}^* = 1$.

The name Data Envelopment Analysis is obtained from the problem on the left in (2) in the following manner. An optimal solution will envelop the outputs of DMU_o from above via constraints of the form $y_{ro} \leq \sum_{j=1}^{n} y_{rj} \lambda_j^*$ with at least one of these r = 1, ..., s constraints satisfied as an equation. Thus there will be at least one "touching" of an observed output for DMU_o by the solution associated with an optimal choice of λ^* values. Similarly, the inputs of DMU_o are enveloped from below via the constraints $\theta_o^* x_{io} \geq \sum_{j=1}^{n} x_{ij} \lambda_j^*$, with at least one of these input constraints satisfied as an equation. Thus, Data Envelopment Analysis is suggested as a name because the output and input data of DMU_o are enveloped from above and below in the indicated manner.

The input and output data in the columns of the problem on the left in (2) are the same as the input and output data in the rows of the problem on the right. It follows that the number of variables in one problem is equal to the number of constraints in the other. The two problems are said to be duals and are connected by a formal mathematical theory which asserts that $y_o \leq h_o$ and, at an optimum, $h_o^* = y_o^* \leq 1$ with $h_o^* = y_o^* = 1$ if and only if DMU_o is fully (i.e., 100%) efficient.

Examination of the problem on the left in (2) shows that full efficiency will be achieved if and only if both of the following conditions are satisfied

$$\begin{aligned} &\theta_o^* = 1 \\ &\text{and} \\ &\text{all slacks are zero.} \end{aligned} \tag{4}$$

As was noted in Sections II and III, the optimal solutions are to be compared to DMU_o's observed performance in arriving at an efficiency evaluation. A value of $\theta_o^* < 1$ in this optimal solution reduces all of the inputs used by DMU_o to a fraction of their observed values. This is interpreted to mean that a combination formed from the data from other DMUs can be located which uses only the amount $\theta_o^* x_{io} < x_{io}$ and this indicates excesses in every input used by DMU_o. If, in addition, any input slack s_i^{-*} is not zero for DMU_o then this slack amount can also be subtracted from the amount of the i[th] input used by DMU_o without altering any other input or output. Thus with nonzero slack for the i[th] input we have $\theta_o^* x_{io} - s_i^{-*} < \theta_o^* x_{io} < x_{io}$. Similarly, a non-zero output slack means a shortfall is present in the output value with which it is associated and there is then an inefficiency in the amounts s_r^{+*} of the r[th] output produced by DMU_o.

The role played by ϵ may be illustrated by returning to Figure 1 and using the following version of the problem on the left in (2) to evaluate DMU_5. For this single input-single output case, the evaluation problem for DMU_5 is secured from Figure 1 for use with (2) in the following form,

$$\begin{aligned} &\min\ \theta_o \qquad -\epsilon\,(s_1^- + s_1^+) \\ &\text{subject to} \\ &0 = \theta_o\, x_{15} - \sum_{j=1}^{5} x_{1j}\,\lambda_j - s_1^- \\ &y_{15} = \sum_{j=1}^{5} y_{1j}\lambda_j - s_1^+ \\ &\lambda_j,\ s_1^-,\ s_1^+ \geq 0, \qquad j = 1, 2, \ldots, 5,. \end{aligned} \tag{5}$$

where x_{15} and y_{15} refer to the amount of input 1 and output 1 used by DMU_5. I.e., we are using the index j, as in (2), to identify which of these 5 DMU_j's is being referred to in (5).

One solution of (5) is $\lambda_5 = 1$ and $\theta_o = 1$ with all other variables equal to zero—including all of the slack variables associated with ϵ in the objective state at the top of (5). However, $\epsilon > 0$ requires the slack to be maximized as stated in th objective for (5) and so the solution $\lambda_4^* = 1$ and $s_1^{-*} = x_{15} - x_{14} > 0$ with the same value of $\theta_o = 1$ improves on the preceding solution. That is, $\theta_o^* - \epsilon s_1^{-*} < \theta = 1$ with $\theta_o^* = \theta$ and $s_1^{-*} > 0$. Hence, the solution with $\theta_o = 1$ and all slacks equal to zero could not have been optimal. It happens that the last solution $\theta_o^* = 1$, $s_1^{-*} > 0$ is optimal and DMU_5 is therefore found to be inefficient even with $\theta_o^* = 1$ because of the excess indicated by this amount of slack in its input. Thus, as in this example, the sum of the slacks is maximized, but this is accomplished by maximizing the slack *without* reducing the optimal value of θ_o^*. Consequently, $h_o^* = y_o^* < 1$ when inefficiency is exhibited for DMU_o by failure of this solution to satisfy either or *both* of the conditions in (4).

It needs to be noted that P_4 was used in evaluating the performance of DMU_5 and, therefore, DMU_4 is technically efficient as shown in Charnes, Cooper and Rhodes (1978). Visual confirmation is obtained by observing that there is no way to move from P_4 to any other point in the set bounded by the solid line in Figure 1 without decreasing output or increasing input. This means that a tradeoff is *necessary* to effect any movement from P_4 to any other point in the "production possibility set." Some kind of pricing or weighting scheme must therefore be used to determine whether any such tradeoff is worthwhile.

What has just been said about efficiency in observed behavior can be formalized as follows:

> 100% relative efficiency is attained by any DMU only when comparisons with other relevant DMUs, which are efficient, do not provide evidence of inefficiency in the use of any input or output. In particular, the conditions in (4) are both satisfied when 100% relative efficiency is achieved and this means that it is not possible to improve some observed input or output value for the thus evaluated DMU_o without worsening other input or output values. (6)

The evaluation of any DMU is effected by reference to some combination of efficient DMUs selected from the set of available points on the efficiency frontier. This is automatically accomplished in DEA computations which can be conducted with any available linear programming code to obtain an optimal solution formed from a relevant set of efficient DMUs.

Generally, it will be desired to evaluate every one of the $j = 1, \ldots, n$ DMUs. This can be accomplished in serial fashion as follows. To evaluate P_1, for example, only the values x_{15} and y_{15} as shown explicitly in (5) would be replaced by x_{11} and y_{11} to obtain an efficiency evaluation of P_1. Then x_{12} and y_{12} would replace x_{11} and y_{11} and the process continued until every one of these $j = 1, \ldots, 5$ DMUs was evaluated. This would be a tedious process to use for a large number of DMUs, of course, but the task can be avoided by recourse to computer codes that have now been designed for use with DEA.[10] Hence a user of these codes is freed for attention to issues other than computation or selection of the pertinent DMUs to be evaluated since this should all be accomplished by the computer code used.

Returning to the case of multiple inputs and outputs, if any inefficiencies are present they may be identified by reference to the conditions in (4). If adjustments to the efficiency frontiers are wanted for any DMU_o, they may be obtained by means of the following "CCR projection formulas":

$$\hat{x}_{io} = \theta_o^* x_{io} - s_i^{-*}, \quad i = 1, \ldots, m$$
$$\hat{y}_{ro} = y_{ro} + s_r^{+*}, \quad r = 1, \ldots, s \tag{7}$$

where the x_{io} and y_{ro} values are originally observed data and the stars indicate optimal values as determined by DEA. Then, as shown in Charnes, Cooper and Rhodes (1978, pp. 433–434), the values of $\hat{x}_{io}$ and $\hat{y}_{ro}$ obtained in this manner will yield an efficient point when these formulas are applied to all $i = 1, \ldots, m$ inputs and $r = 1, \ldots, s$ outputs for any DMU_o.

Evidently, n optimizations, one for each of n DMUs, are used in a complete DEA analysis.[11] For each optimization, the computer printouts should make information available on the efficient DMUs (mathematically speaking, this is part of the basis set) from which the inefficiencies were derived since these may be needed to guide interpretations and the tests and extensions that should be investigated.

Examples are provided in the articles that follow in the form of standard computer printouts. See, e.g., Exhibits 8 and 9 in the following article. The examples in that article use the problem on the left in (2). However, either of the dual pair of problems in (2) may be used, but if the problem on the right is used, the two conditions for efficiency in (4) give way to the following single condition:

$$y_o^* = \sum_{r=1}^{s} \mu_r^* y_{ro} = 1. \tag{8}$$

Achievement of this condition for 100% efficiency implies that no ϵ enters into any of the optimal μ_r^* values.[12]

The above dual pair of linear programming problems represented in (2) were originally published in Charnes, Cooper and Rhodes (1978) to make it possible to obtain solutions for the following problem:

$$\max \hat{h}_o = \sum_{r=1}^{s} u_r y_{ro} \Big/ \sum_{i=1}^{m} v_i x_{io}$$

subject to

$$\sum_{r=1}^{s} u_r y_{rj} \Big/ \sum_{i=1}^{m} v_i x_{ij} \leq 1, \; j = 1, \ldots, n \tag{9}$$

$$u_r \Big/ \sum_{i=1}^{m} v_i x_{io} \geq \epsilon, \; r = 1, \ldots, s$$

$$v_i \Big/ \sum_{i=1}^{m} v_i x_{io} \geq \epsilon, \; i = 1, \ldots, m$$

The reason for the name CCR ratio form should now be evident. All constraints and the objective are stated in ratio form. In this form the problem is interpretable as follows: Subject to these constraints choose the u_r and v_i values that will assign the maximum possible ratio value, $\hat{h}_o^*$, which is to be used to rate DMU_o's efficiency.

This is a nonlinear problem that would give rise to computational difficulties in trying to obtain the solutions required for practical use of DEA. However, the following version of the CC transformation from the theory of fractional programming, as first given in Charnes and Cooper (1962), can be used to replace (9) with (2).

$$u_r = t\mu_r, \; r = 1, \ldots, s$$
$$v_i = t\nu_i, \; i = 1, \ldots, m \tag{10}$$
$$t = 1 \Big/ \sum_{i=1}^{m} \nu_i x_{io}$$

Nothing is lost. Since all variables are positive, it is also possible to transform from (2) to (9) by means of these same formulas. Furthermore, by proceeding in this way, contact is made with the theory of fractional programming as given in Charnes and Cooper (1962) so that also

$$\max \hat{h}_o = \hat{h}_o^* = y_o^* = h_o^* \tag{11}$$

Hence, all three problems are available for use in computation or interpretation for Data Envelopment Analysis. Also, as noted in Charnes, Cooper and Rhodes (1978), still other problems and model types can be introduced via this route and this includes those that are related to each other via the duality theory of fractional programming. See, e.g., Schaible (1974).

Thus, via the transformations given in (10), access is provided to the computationally efficient forms given in (2). Either of these dual pair of linear programming problems may be used. Moreover, an optimal solution to the thus selected member of (2) also provides access to the optimal solutions for the other two problems as well. The analytical and interpretive power available from (9), can thus be added to the interpretations and alternative modeling possibilities which are available from the dual pair in (2).

It perhaps suffices to indicate what is analytically added by noting that (9) can be used to supply a proof, as in Charnes and Cooper (1985), that the optimal value of $\bar{h}_o^*$ is invariant to changes in the units used to measure any input or output. That is, a change in the units used to measure any input or output—such as a change from pounds to tons—will not alter the value of max $\bar{h}_o = \bar{h}_o^*$. The relation in (11) means that this invariance property is also applicable to both of the problems in (2). Thus the units used for any input or output may be chosen as desired for *any* of these models without affecting these optimum values.[13]

As already indicated, the model in (9) is what gave rise to the name "CCR ratio form" which, via the use of the theory of fractional programming, is transferable to the dual pair of linear programming problems in (2). Using the concepts of virtual output and virtual input as discussed in connection with (2), it is possible to make contact with the usual one-output-to-one-input ratio measures of efficiency used in engineering and the natural sciences. Conversely, DEA can be regarded as generalizing these engineering and natural science measures for use with multiple outputs and multiple inputs and this is also done in a manner that brings these concepts from engineering and science into contact with concepts in economics (such as the concept of Pareto-Koopmans efficiency). See Charnes, Cooper and Rhodes (1978, pp. 430–431).

This virtual output-to-virtual input interpretation also provides access to easily usable and understandable graphical portrayals. These portrayals like the one in Figure 1 are available if its y and x values are regarded as having been derived from (9) so that, in particular, $\bar{h}_o^* = y_o^*/x_o^* \le 1$ with $y_o^*/x_o^* = 1$ if and only if the DMU_o being evaluated has achieved full (100%) efficiency.[14]

V. BCC MODELS AND RETURNS TO SCALE

The dual problems represented in (2) are now replaced by the following new pair:

$$
\begin{array}{ll}
\min\ \theta_o - \epsilon\,[\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+] & \max\ \sum_{r=1}^{s} \mu_r y_{ro} - u_o \\
\text{subject to} & \text{subject to} \qquad\qquad (12) \\
0 = \theta_o x_{io} - \sum_{j=1}^{n} x_{ij}\lambda_j - s_i^- & \sum_{i=1}^{m} \nu_i x_{io} = 1 \\
y_{ro} = \sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ & \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \nu_i x_{ij} - u_o \le 0 \\
1 = \sum_{j=1}^{n} \lambda_j & \mu_r \ge \epsilon \\
 & \nu_i \ge \epsilon \\
\end{array}
$$

$$0 \le \lambda_j, s_i^-, s_r^+ \quad \text{for } i = 1, \ldots, m;\ r = 1, \ldots, s;\ j = 1, \ldots, n.$$

Comparison will show that this differs from (2) only in that the new $\sum_{j=1}^{n} \lambda_j = 1$ condition appears at the bottom of the problem on the left and a new variable u_o appears in the problem on the right. The correspondence between the number of variables in one problem and the number of constraints in the other is thus maintained. Via this correspondence the variables in one problem can be used to evaluate a constraint in the other problem in terms of its effects on the objective being pursued. See Chapter I in Charnes and Cooper (1961) for a discussion of some of the managerial uses of these duality relations.

One reason for introducing the condition $\sum_{j=1}^{n} \lambda_j = 1$ in (12) is to ensure that all solutions and, hence, their associated DMU efficiencies will be evaluated only by reference to original data points and their "convex combinations"—i.e., by percentage combinations which add to 100 percent. These combinations can then be used to generate comparison points on efficiency frontiers. Compare the discussion of such points formed from P_3 and P_4 in Figure 1 as potential candidates for evaluating DMU_1. Note, in particular, that points on the ray stretching beyond P_3 on the dotted line in Figure 1 are thus eliminated as possible candidates for evaluating points like P_1 by virtue of this "convexity condition."

The new variable u_o which appears in the right-hand problem of (12) is now available for identifying returns to scale possibilities in accordance with the following criteria:

$$
\begin{array}{l}
u_o^* < 0 \Rightarrow \text{increasing return to scale} \\
u_o^* = 0 \Rightarrow \text{constant return to scale} \\
u_o^* > 0 \Rightarrow \text{nonincreasing return to scale,}
\end{array}
\qquad (13)
$$

where the stars indicate that this value of u_o is part of an optimum solution and the symbol "⇒" means "implies."

As noted in Banker, Charnes and Cooper (1984) where these properties are developed analytically on the assumption of a unique optimum, the following two points need to be born in mind:

First, the thus indicated returns to scale are "local" in that they are applicable only to the facet on the efficiency frontier where the reference point for the efficiency evaluation is positioned. (Refer to Figure 1, for example, where the facet from P_3 to P_4 displays returns which are locally decreasing while the facet from P_2 to P_3 displays returns which are locally increasing while the returns at P_3 are locally constant.)

Second, in actual use the possibility of alternate optima also needs to be considered. For the general case of m inputs and s outputs, this topic is rigorously addressed in Banker and Thrall (1988) with results that we summarize as follows: Local increasing returns are said to prevail if and only if $u_o^* < 0$ for all alternate optima. Local decreasing returns to scale are said to prevail if and only if $u_o^* > 0$ for all alternate optima. All of the remaining cases correspond to local constant returns to scale.

Using the concept of Most Productive Scale Size, it is shown that the pertinent information can also be extracted from the solutions to (2) in ways that relate them to sign conditions in (13) even though the extra variable u_0 does not appear in these models. See also Banker and Thrall (1988) and Charnes and Cooper (1985).

$$\max \frac{\sum_{r=1}^{s} u_r y_{ro} - u_o}{\sum_{i=1}^{m} v_i x_{io}}$$

subject to

$$1 \geq \frac{\sum_{r=1}^{s} u_r y_{rj} - u_o}{\sum_{i=1}^{m} v_i x_{ij}} \qquad (14)$$

$$\epsilon \leq u_r \Big/ \sum_{i=1}^{m} v_i x_{io}$$

$$\epsilon \leq v_i \Big/ \sum_{i=1}^{m} v_i x_{io}$$

for $j = 1, \ldots, n$; $r = 1, \ldots, s$; $i = 1, \ldots, m$.

Inspection shows that this is the same as (9) except for the appearance of u_0 in (14). In fact, the transformation in (10) that was used to go from (2) to (9) can be used to relate (14) to (12) and so no further discussion is needed to describe relations between the BCC ratio form in (14) and their linear programming equivalents in (12).

VI. ADDITIVE MODELS

As noted in Section II, it is sometimes desirable to test results by using different models. Hence, we introduce another class of models called the "additive models" as introduced in Charnes, et al. (1985) which also incorporates the concepts and methods of DEA. These models are of interest in their own right, and lead to still other uses and extensions of DEA. They can be formalized as follows:

$$\max \sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^- \qquad\qquad \min \sum_{i=1}^{m} \nu_i x_{io} - \sum_{r=1}^{s} \mu_r y_{ro} + u_o$$

subject to: $\qquad\qquad$ subject to: $\qquad$ (15)

$$-y_{ro} = \sum_{j=1}^{n} y_{rj}\lambda_j + s_r^+ \qquad\qquad \sum_{r=1}^{m} \nu_i x_{ij} - \sum \mu_r y_{rj} + u_o \geq 0$$

$$x_{io} = \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- \qquad\qquad \nu_i \geq 1$$

$$\mu_r \geq 1$$

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

$$0 \leq \lambda_j,\ s_r^+,\ s_i^- \text{ for } i = 1, \ldots, m;\ r = 1, \ldots, s;\ j = 1, \ldots, n.$$

Starting with the problem on the left, the only test for efficiency is whether all slacks are zero. That is, DMU_o is fully efficient if and only if

$$\max \sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^- = \sum_{r=1}^{s} s_r^{+*} + \sum_{i=1}^{m} s_i^{-*} = 0 \qquad (16)$$

As can be seen, there are no ϵ in either of the above problems. Hence, such non-Archimedean elements need not be considered.

Figure 2 can help to show differences in the way the additive and the ratio models locate points from which inefficiencies may be determined. For this Figure, there are two inputs to be considered in amounts x_1 and x_2 and only one output y. Division of each of the inputs by the output amount with which it is associated converts these into rates of input per unit output. This is why $\hat{x}_1$ and $\hat{x}_2$ replace x_1 and x_2 as coordinates in their respective dimensions in Figure 2.

Given the data plotted in Figure 2, it is desired to evaluate the performance of DMU_5 with the input rates $(\hat{x}_{15}, \hat{x}_{25})$. This is to be accomplished by selecting a relevant point on the efficiency frontier. Here the efficiency frontier is represented by the "unit isoquant" as is shown by the solid line connecting P_2 and P_3. A ratio model determines this point via a value θ_o^* which can be interpreted in terms of the ray indicated by the broken line from the origin to P_5. In fact, the

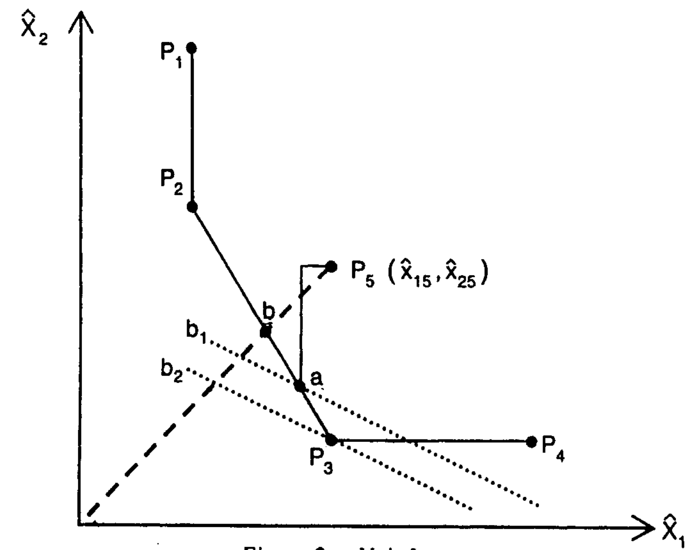

Figure 2. Unit Isoquant

optimal value of θ_o* is geometrically interpreted as the length of the ray from the origin to b divided by the length of the origin to P_5. Thus

$$0 \leq \theta_o^* = \frac{\ell(0,b)}{\ell(0,P_5)} \leq 1 \qquad (17)$$

where $\ell(0,b)$ refers to the length of the ray represented by the portion of the broken line from the origin to b and $\ell(0,P_5)$ refers to the length from the origin to P_5.[15]

The additive model, on the other hand, selects a point such as a, which maximizes the sum of the slacks, $s_1 + s_2$, where the value of s_1 is geometrically represented by the length of the solid horizontal line emanating from P_5 and the value of s_2 is represented by the length of the solid vertical line extending to the frontier.[16] Although the amounts of inefficiencies designated by these two models will differ because different points on the frontier are thus selected, it is nevertheless important to note that both models characterize DMU_5 as inefficient.

It is not true that the amounts of the different inefficiencies will always differ when they are obtained by these two different types of models. Notice, for instance, that the points P_1 and P_4 which are on the frontier in Figure 2 are not

efficient.[17] The additive and ratio models will not only characterize them as inefficient but they will also do so in the same way–namely, by maximizing the slack value of s_1 which is associated with the first input for P_4 and by maximizing the slack value of s_2 which is associated with the excess in the second input used by DMU_1. Finally, P_2 and P_3 will both be identified as efficient (with zero slacks) by either the additive or ratio approaches. See Ahn, Charnes and Cooper (forthcoming, 1989) for an analytical development of relations between these models.

Turning to the problem on the right in (15) we note that the condition that every input and every output has some positive value is now reflected in the constraints which require $\nu_i \geq 1$ all i and $\mu_r \geq 1$ all r. These can be regarded as normalizations imposed on the values of these variables—which are assigned to the similarly indexed inputs and outputs in the objective stated for the problem on the right in (15).

These input and output values in amounts x_{io} and y_{ro} in the objective of the problem on the right are dually related to the stipulations for the constraints in the problem on the left in (15). Hence, we may interpret the objective for the problem on the right as being oriented toward minimizing the "efficiency losses" occasioned by the deviations from the observed values to which they are related.

The objective of the problem on the right can be reoriented to a gain, or profit maximizing, objective, as in the following:

$$\max \sum_{r=1}^{s} \mu_r y_{ro} - \sum_{i=1}^{m} \nu_i x_{ij} - u_o. \qquad (18)$$

Using stars to indicate optimal values, we employ the duality theory of linear programming which relates this objective to the one stated in (16) by asserting that their values will be equal at an optimum. Thus, we also have

$$\sum_{r=1}^{s} \mu_r^* y_{ro} - \sum_{i=1}^{m} \nu_i^* x_{ij} - u_o^* = 0 \qquad (19)$$

if and only if DMU_o is efficient.

The optimal values μ_r* and ν_i* may be used to evaluate the effects of increments or decrements in the outputs and inputs with which they are associated. These evaluators can be interpreted as providing "efficiency prices" for evaluating the rates at which tradeoffs may be affected on the frontiers with which these values are associated.[18] These tradeoffs are determined by reference to ratios of these dual variables but they need not involve any monetary units (e.g., dollars) and so terms like "profits" and "losses" and "prices" are to be regarded as only generic usages in the above discussions.

As was remarked earlier, it is necessary to introduce external prices, or to use relative weights in order to determine whether such tradeoffs should be undertaken. The dotted lines in Figure 2 provide an example which can also be used to illustrate the concept of "allocative efficiency." Theoretical considerations in the evaluations of allocative efficiency using the BCC model may be found in the article by Banker and Morey in this issue. See also Banker (1988) and Banker and Maindiratta (1988).

Here we develop the concept of allocative efficiency in a simplified way by assuming a price p_1 for x_1, and a price p_2 for x_2 so that $p_1 x_1 + p_2 x_2$ represents a total cost of using these amounts of x_1 and x_2.[19] These costs can be associated with a "budget line" such as the one portrayed by the dotted lines in Figure 2. For instance, $b_2 = p_1 \hat{x}_{13} + p_2 x_{23}$ uses the coordinates of P_3 to obtain the total cost of producing a unit output with these inputs. This is lower than the cost b_1 associated with the inputs at a. Hence, a is technically but not allocatively efficient. Thus, if cost minimization is an objective, the tradeoffs involved in moving from a to P_3 are worth undertaking with these prices.

Methods for determining amounts of allocative as well as technical efficiency have been available since the early work of Farrell (1957) but they all depend on an exact knowledge of prices to be considered. See Färe, Grosskopf and Lovell (1985) for discussion of these methods. Only recently has it become possible to relax the requirements for an exact knowledge of prices so that it is now possible to proceed with a much looser requirement in which only ranges (in the form of upper and/or lower bounds) on the possible prices or weights are required. This looser approach is the one which is most likely to be of use in evaluating governmental and nonprofit activities. See Banker and Morey, this issue, or Charnes, Cooper, Huang and Sun (forthcoming, 1989) and Thompson, Singleton, Thrall and Smith (1986). See also Thrall (forthcoming, 1989).

The variable u_o in the right-hand problem of (15), which remains to be explained, plays the same role here as was displayed in (13) for the BCC model. We can show this as a byproduct of the following development which also characterizes relations between the additive and the BCC models (and hence also to the CCR models).

Reference to the constraints in the right-hand problem of (15) shows that they can also be written

$$1 \geq \frac{\sum_{r=1}^{s} \mu_r y_{rj} - u_o}{\sum_{i=1}^{m} \nu_i x_{ij}} \quad j = 1, \ldots, n, \tag{20}$$

where this division by $\sum_{i=1}^{m} \nu_i x_{ij}$ does not alter the sense of the inequality because positive values for the denominator are guaranteed by the condition $\nu_i \geq 1$ for all i, which every solution must satisfy. Comparison shows that the constraints in (20) are in the same form as the ones displayed in (14). Using the transformations in (10) brings the two together in all respects with the ϵ conditions for positivity also being satisfied *a fortiori* by any solution to (20) which is subjected to this transformation.

From the duality relations of linear programming, which are applicable to the dual pair in (15), we have

$$0 \leq \sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^- \leq \sum_{r=1}^{m} \nu_i x_{io} - \sum_{r=1}^{s} \mu_r y_{ro} + u_o \tag{21}$$

with equality holding between the solution values of these two problems only for an optimal pair. The expression on the right of (21) therefore always satisfies

$$\frac{\sum_{r=1}^{s} \mu_r y_{ro} - u_o}{\sum_{i=1}^{m} \nu_i x_{io}} \leq 1. \tag{22}$$

Noting that this is in the same form as the functional used in the objective of (14), we can again proceed via the transformations in (10) to extend this correspondence by identifying these results with the same variables as in (14). This is accomplished without altering the limit to solutions on the right in (22)—which is also the same limit value for the maxima attainable in (14).

To complete this development, we observe that the optimal choices of s_r^+, s_i^- are not altered if these terms are all multiplied by $\epsilon > 0$ in the objective stated for the left-hand problem in (15). This multiplication replaces the conditions $\nu_i, \mu_r \geq 1$ with $\nu_i, \mu_r \geq \epsilon$, all i and r, in the problem on the right so that, combining this result with (20) and (22), we obtain the model depicted in (14) in all detail. It follows that a DMU_o found to be efficient in (15) will also be found to be efficient in (12) and vice versa. The same holds for inefficiency: i.e., DMU_o will be characterized as inefficient by the BCC ratio form if and only if it is characterized as inefficient by the additive form. See Ahn, Charnes and Cooper (1989) for a fuller development of these and other relations between different DEA models.

Various byproducts of the thus demonstrated relations are available if wanted. It may be desired, for example, to employ ratio efficiency values rather than sums of the slack deviations from efficiency. Reference to (22) shows how this may be done *after* an optimum has been achieved with full efficiency being obtained if and only if the resulting ratio is equal to unity.

Finally, contacts with other disciplines provide further opportunities for both use and research. For instance, as demonstrated in Charnes, Cooper, Golany, Seiford and Stutz (1985) these additive models are directly related to goal programming in a form that also gives direct access to the activity analysis models of economic theory as described in Chapter IX of Charnes and Cooper (1961). Therefore, as described in Bowlin (1984), results from such efficiency analyses can be extended beyond the evaluation of individual DMUs by means of goal programming models in which these efficiency adjusted results can then be used for budgetary purposes in which resource reallocations between efficient DMUs may be used to improve the performance of a total system.

VII. EXTENDED ADDITIVE MODELS

In many applications, it may be necessary to consider inputs and outputs which are exogenously fixed and hence cannot be varied at the discretion of the managers of different DMUs. For example, the number of successfully completed sorties represents an output which entered into the efficiency evaluations of the maintenance activities in U.S. Air Force wings as described in Charnes, Clark, Cooper and Golany (1985). The possible number of successful sorties is evidently dependent on weather. Thus weather, although a variable beyond the control of management, must nevertheless be taken into account in effecting efficiency evaluations in bases that are located in places that are as different as southwest Texas and northern Alaska.

Banker and Morey (1986) have provided one way of taking nondiscretionary variables like weather into account in BCC and CCR ratio models. Here we shall provide another recently developed approach, as given in Charnes, Cooper, Rousseau and Semple (1987), for use with additive models. This approach, as will be seen, also extends the class of additive models in other ways as well.

In addition to preserving the conditions for Pareto-Koopmans efficiency[20] for use in their evaluations, these "extended additive models" also make it possible to deal with thresholds and ceilings when they need to be taken into account. As an illustration of thresholds and ceilings that can bear on modeling for managerial use we turn first to a statistical model used in a Department of Defense sponsored study which recommended reducing the budgeted amount of Marine Corps recruitment advertising. The resulting budget would have fallen below the threshold that is needed for *any* national TV advertising to be undertaken. The unavailability of any TV advertising, in turn, would have made it necessary for the Marine Corps to revise other parts of its recruiting activities in ways that were not reflected in the statistical model that had been employed.[21] See Charnes, Cooper and Golany (1986a, 1986b) for a critique.[22]

Ceilings may also need to be considered. For instance, it is a practice in the Tactical Air Command to conserve fuel by halting further sorties of fighter aircraft at an Air Force base when a target number of sorties is attained. This needs to be taken into account in efficiency evaluations, while also allowing for continuance of maintenance activities and the training of mechanics which are also important outputs. See Charnes, Clark, Cooper and Golany (1985).

All of these considerations, and more, are taken into account explicitly in the following extended additive model:

$$\max \sum_{r=1}^{s} \frac{s_r^+}{|y_{ro}|} + \sum_{i=1}^{m} \frac{s_i^-}{|x_{io}|}$$

subject to

$$\begin{aligned} -y_{ro} &= -\sum_{r=1}^{s} y_{rj}\lambda_j + s_r^+ \\ x_{io} &= \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- \\ 1 &= \sum_{j=1}^{n} \lambda_j \\ \gamma_r y_{ro} &\geq s_r^+ \\ \beta_i x_{io} &\geq s_i^- \end{aligned} \tag{23}$$

$$0 \leq s_i^-, s_r^+, \lambda_j; \text{ for } i = 1, \ldots, m;\ r = 1, \ldots, s;\ j = 1, \ldots, n.$$

The β_i parameter values are fixed by the modeler somewhere in the range of zero to one in accordance with considerations that may be clarified by writing the condition for the ith constraint as follows:

$$x_{io} = \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- \leq \sum_{j=1}^{n} x_{ij}\lambda_j + \beta_i x_{io}. \tag{24}$$

Now for $\beta_i = 0$ these expressions will be satisfied only for $s_i^- = 0$. This condition must also hold at an optimum in which case we have $x_{io} = \sum_{j=1}^{n} x_{ij}\lambda_j^*$ and no penalty is incurred for the thus indicated consumption of x_{io} which was entirely non-discretionary. Turning to $\beta_i = 1$ the completely discretionary case comes into play with any positive slack values $s_i^{-*} = x_{io} - \sum_{j=1}^{n} x_{ij} \lambda_j^*$ to be reflected in the objective if it is part of an optimum solution.

Finally, for $0 < \beta_i < 1$ we can rewrite (24) as

$$(1 - \beta_i)\, x_{io} \leq \sum_{j=1}^{n} x_{ij} \lambda_j \tag{25}$$

which shows that the combination on the right must equal or exceed the threshold value on the left. Hence, the threshold on the left is taken into account by the expression on the right when it designates the efficient amount of the i^{th} input.

Similar reasoning applies to the γ_r choices. These range from 0 to 1. For each such γ_r choice we obtain for the r^{th} output,

$$y_{ro} = \sum_{j=1}^{n} y_{rj} \lambda_j - s_r^+ \geq \sum_{j=1}^{n} y_{rj} \lambda_j - \gamma_r\, y_{ro}. \tag{26}$$

The reasoning is analogous to the input case. Allowance being made for subtraction rather than the addition of positive slack, the reversal of the inequality in (26) changes the threshold characterization into a ceiling for the outputs to which the inequality $\gamma_r\, y_{ro} \geq s_r^+$ is applied.

Finally, we turn to the function used for the objective in (23) where the vertical strokes on the terms in the denominator mean that an absolute (positive) value is to be used. This implies that the maximization is oriented toward *relative* goal deviations and, because of the way the goals are stated in the denominator, the resulting value is independent of the units of measurement used.[23]

VIII. CONCLUSION

The different types of DEA models that have been presented do not exhaust this topic. Others are also available–such as the multiplicative models presented in Charnes, Cooper, Seiford and Stutz (1983)—which can be regarded either as providing alternatives to those discussed in the preceding sections or as providing new possibilities in which combinations of different DEA models can be used. See Banker and Maindiratta (1986) for an example of how multiplicative and BCC models may be joined.

DEA continues to be an active area of research in its own right so that still other possibilities are likely to become available. It thus seems better to turn to topics dealing with actual uses of DEA in place of further development of underlying concepts and available DEA models. The three following papers provide examples that can help to supply possible study strategies.

When actual managerial uses are a study objective it is generally advisable to start with the kinds of information that management customarily uses in its evaluations. This can help to facilitate managerial understanding of study results and it can also make it easy for managers to participate in initial formulations and subsequent evaluations of study results.

One way this can be accomplished is illustrated in the second paper in the following collection which deals with a study of possible uses of DEA in evaluating the performance of electric cooperatives in Texas. Much of the work underlying this study was taken from Thomas (1986), in which, as a member of the Texas Public Utility Commission (and subsequently its chairman) Dr. Thomas was very much interested in improved approaches that could be used to evaluate utility performance. One reason for this interest is that a relatively recent statute in Texas requires management (efficiency) audits to be performed periodically for electric cooperatives in Texas. This statutory requirement helped to guide the study choices of inputs and outputs as well as choices of the tests that might be used for evaluating DEA. *Inter alia*, this led to comparisons with results from field audits that had been conducted at individual DMU (=individual electric cooperative) levels as well as comparisons with other evaluative devices such as the REA (Rural Electrification Administration) ratios and regressions that were discussed in the opening two sections of our paper.

The need for guidance in allocating audit resources also led to the development of a suitable means for obtaining a ranking of the 75 electric cooperatives in Texas for Commission use in directing its field audits. This was accomplished, as indicated in the article, by providing a transformation of all inefficiencies, including lost revenue opportunities, into dollar terms. This way of effecting the rankings makes it possible to deal with slack and scale as well as technical inefficiencies and can also make allowances for the different comparison sets that may enter into these efficiency evaluations. These developments in DEA therefore illustrate the use of managerial know-how not only in selecting data but also in providing openings into new problem areas and possible new uses of DEA that might otherwise have been overlooked by adherence to an initial study design.

The first paper in the series that follows provides an example of a different strategy which was accomplished in two different stages as follows. The first stage took advantage of a study that was already under way by staff of the Select Committee on Higher Education (SCOHE) which had been tasked by the Texas legislature to study the performances of state institutions of higher learning (IHLs) in Texas. Members of this staff kindly supplied help and advice in selecting inputs and outputs (as well as the choices of DMUs) so that results from DEA could be readily compared with those obtained by Commission staff in studies which were employing other methods of analysis. See Ahn (1987) for further discussion and acknowledgments. A second stage then followed in which Victor Arnold was invited to review these DEA results. The opportunity provided by Dr. Arnold's completion of his duties as Director of SCOHE was also used to invite Dr. Arnold to bring his already acquired know-how to bear in the endeavors that led to the paper that now appears in the following collection—which reports on the use of DEA in evaluating the performance of state IHLs in Texas.[24]

The third paper in the following series, in which Richard Clarke appears as a coauthor, provides an example of yet another approach which took advantage of Dr. Clarke's experience as a functioning U.S. Air Force officer, as well as his responsibilities for research in logistics when he was on the faculty at the U.S. Air Force Institute of Technology at the Wright Patterson AFB. This paper is part of a larger study which deals with vehicle maintenance activities in the Tactical Air Command of the U.S. Air Force. In this study, Dr. Clarke was able to take advantage of an opportunity that was provided by an Inspector General's report which criticized the efficiency measures that were then being employed. In the example reported here, the orientation is more toward "scientific testing" rather than "actual use" in that this part of the study is directed to providing a new approach for testing for the presence of "organizational slack"—a concept that plays a prominent role in the modern literature of organization theory. The discussion and developments reported here are, however, only part of a larger effort reported in Clarke (1988) in which issues of user attitudes and active managerial participation are also addressed as part of a study on the effects of repeated uses of DEA on managerial performance.

Topics of modeling and testing, as well as opportunities for research and use, need to be considered from new standpoints as DEA and its uses are still being developed. Familiarity with underlying concepts is advisable along with a knowledge of available alternatives (e.g., alternative models) and their uses and interpretations. The same is true of testing and validation procedures.

We have already indicated how managerial uses (including field audits) can be brought to bear in testing and validation. Strong arguments can be made for exploiting managerial know-how and experience in interpreting results at individual DMU levels, as well as initial and subsequent formulations, especially when a use of DEA as a managerial tool represents a study objective. Thus, after comparing DEA results with results from the ratio and other analyses used by SCOHE, the first paper in the following set concludes with a "window analysis."[25] As yet another example of new approaches to testing, such "window analyses" can be used to provide a systematic check of stability of the efficiency evaluations over different collections of DMUs while also increasing the number of degrees of freedom. In addition to providing added protection against degrees of freedom deficiencies, such an analysis can be regarded as providing trend information which can be identified with the behavior of individual DMUs in a form that can also be submitted for review and readily interpreted by managers familiar with the performance of these DMUs.

The uses of comparisons with field audits at individual DMUs was noted earlier as a way of checking the results of a DEA study of electric cooperatives in Texas in the second of the following three papers. In addition, correlations between DEA and the results of REA regression and ratio analyses were undertaken in order to see whether (or in what way) DEA might confirm or conflict with results from these other analytical approaches.

This kind of comparative use of DEA and statistical regressions does not exhaust the possibilities for their joint uses. For instance, DEA and statistical regression approaches might be joined, as was done by Rhodes and Southwick (1986) in their study of the relative efficiency of private and public universities. They did this by first applying formula (7) to project all data onto efficiency frontiers before undertaking regression estimates from the thus adjusted data. In addition to providing new types of uses for regressions by fitting them to efficiency adjusted data, this approach makes it possible to conform to the efficiency assumptions of economic theory when studying phenomena such as are associated with returns to scale.

In a very different way, the paper on vehicle maintenance, which is the third of the following three papers, shows how DEA can be used to identify and repair lacunae in the existing literatures of other disciplines. Thus, using a game theoretic interpretation of DEA due to Banker (1980), it was possible to notice a lacuna in the organization theory literature in that the strategies that might be used by managers for accumulating organization slack had been inadequately attended to. As explained in the third of the following papers, this was repairable by reinterpreting a conjecture by Banker so that it was identifiable with points like P_1 and P_4 in Figure 2 where, quite literally, the managers of Air Force vehicle maintenance units could accumulate slack while giving an appearance of fully efficient behavior.

No discussions of these game theoretic formulations and their potential uses for DEA were provided in the preceding sections of this paper because (a) references are available as cited in the third paper in the following series, and (b) pursuit of this topic would have lengthened the present paper and carried it into numerous deviations. Suffice it to say that Banker's game theory formulation of DEA can be shown to be equivalent to the dual pair of linear programming problems in (2) without any of the ϵ values. Although this represents a deficiency from the standpoint of a DEA analysis, it is of interest that even this deficiency was useful in providing opportunities that would not have otherwise been available in testing for the presence of slack—since a use of the ϵ elements precludes any possibility for P_1 or P_4 of Figure 2 appearing to be efficient. At any rate, this third paper in the following set shows how DEA can be used not only to test but also to locate deficiencies in the current scientific literature—and this includes the possibility of empirically testing game theoretic assumptions such as the assumption that optimal strategy choices will always be employed. This and the other papers that follow can suggest still other such possibilities.

Although the emphases in the following papers are on management and accounting (including audit) uses, further opportunities may still be opened by remembering that new principles of estimation are being suggested and these principles may be used in place of or in conjunction with currently used methods in statistics. With DEA, a vehicle is also being provided for extending and adapting cost and production theory as constructs from economics as well as from programming theory and game theory in management science. As noted in the text, efficiency concepts in engineering and the natural sciences are extended and joined to those in economics via DEA, and this suggests that the two may also be joined for use in a variety of applications that might otherwise be beyond the scope of either alone.

This is as far as the present paper will go but the references that follow can suggest additional possibilities for those who want to pursue them. Recent work by Banker and his associates have provided a basis for still further possible relations in which econometric theory with accompanying statistical tests can be brought to bear for use in interpreting DEA. An overview provided by Banker appears below, in this same issue, and it is followed by applications of these concepts in efficiency evaluations and variance analyses.

ACKNOWLEDGMENTS

The research for this paper was partly supported by National Science Foundation Grant SES 8520806 and U.S. Army Contract DAKF-15-87-C-0110 with the Center for Cybernetic Studies at The University of Texas. It was also partly supported by the IC² Institute of The University of Texas at Austin. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

NOTES

1. The task is analogous to what is accomplished in industrial standard costing practices except that DEA extends its identification of inefficiencies to outputs as well as inputs and it does not require recourse to the time, motion and method studies of industrial engineering.
2. See Charnes, Cooper and Rhodes (1978) for a discussion of what can be done when values from theoretically known efficient technologies are available.
3. See the definition of evidence in sense 2, as given by *Kohler's Dictionary for Accountants, 6th ed.* (in Cooper and Ijiri, 1983).
4. Comparisons with DEA could also have been employed for this purpose. See Banker, Charnes, Cooper and Maindiratta (1987) for a comparison of such regressions with a use of DEA in obtaining efficiency evaluations. See also Banker, Conrad and Strauss (1986). Färe, Grosskopf and Lovell (1985) provide comprehensive discussions of various approaches to frontier estimation.
5. A call for papers for an NSF-IC² Institute sponsored conference to be devoted to new managerial uses of DEA (September 28–29, 1989, in Austin, Texas) resulted in receipt of nearly 60 abstracts submitted from different countries in a period of only a few months with many of them dealing with new or underresearched topics.
6. Reference is to the "additive" and "CCR" and "BCC ratio forms" of DEA described in the sections that follow but not to the multiplicative models and their corresponding ratio forms described in Charnes, Cooper, Seiford and Stutz (1983) or Banker and Maindiratta (1986).

7. A fully rigorous theoretical development is provided in Banker and Thrall (1988).
8. The general theory for dealing with non-negative inputs and outputs is given in Charnes, Cooper and Thrall (1986, 1988).
9. ϵ may be interpreted as the reciprocal of the "big M" that is often used in association with artificial variables in linear programming problems. See, e.g., pp. 174 ff. in Charnes and Cooper (1961) which includes a discussion of how these values—which are referred to as "non-Archimedean" values in mathematics—may be treated in simplex tableaux without requiring their values to be specified numerically.
10. Such a code is available from the Center for Cybernetic Studies in The University of Texas at Austin, Texas.
11. In actual use, the computer codes used in DEA take advantage of prior solutions in ways that reduce the number of optimizations actually used.
12. This forms part of the so-called "non-Archimedean theorem" in DEA as given in Charnes and Cooper (1985) in that, via the duality theory of linear programming, it follows that all slacks must be zero and $\theta_o^* = 1$ in the problem on the left in Eq. (2).
13. The non-Archimedean ϵ elements in (9) play the same role as was discussed in connection with (2).
14. This test for efficiency—really "Pareto" or "vector" optimality—was first given in Charnes and Cooper (1957). See also Ch. IX in Charnes and Cooper (1961).
15. These lengths are stated in the ℓ_2 (or Euclidean) metric. See Appendix A in Charnes and Cooper (1961).
16. These lengths are stated in the ℓ_1 metric described in Appendix A of Charnes and Cooper (1961).
17. Recall the discussion of P_5 in Figure 1.
18. See the discussion of these efficiency prices and their uses in Chapter IX of Charnes and Cooper (1961).
19. To be technically correct, these prices need to be stated relative to the output which serves as a divisor of x_1 and x_2.
20. See the discussions of (6) and (7).
21. See the contrast between "control prediction" and the other types of predictions that are described in Charnes, Cooper, Learner and Phillips (1985).
22. A two-stage DEA model for treating advertising as an output in a first-stage, which then becomes an input to a second stage, may be found in Charnes, Cooper, Golany, Halek, Schmitz and Thomas (1986).
23. See Charnes, Cooper, Golany, Seiford and Stutz (1985) or Bowlin, Brennan, Charnes, Cooper and Sueyoshi (1984).
24. Another, more academically oriented DEA study of all U.S. Institutions of Higher Learning (both private and public) is given in Ahn, Charnes and Cooper (forthcoming, 1989b). See also Ahn (1987).
25. This type of analysis was first reported in Charnes, Cooper, Divine, Klopp and Stutz (1982).

REFERENCES

Afriat, S.N., "Efficiency Estimation of Production Functions," *International Economic Review*, XIII, 1972.

Ahn, T., "Efficiency and Related Issues in Higher Education: A Data Envelopment Analysis Approach," Ph.D. Thesis (Graduate School of Business, The University of Texas at Austin, 1987). Also available from University Microfilms International, Ann Arbor, Michigan.

Ahn, T., A. Charnes and W.W. Cooper, "A Note on the Efficiency Characterizations Obtained in Different DEA Models," *Socio-Economic Planning Sciences* (forthcoming, 1989a).

Ahn, T., A. Charnes and W.W. Cooper, "Some Statistical and DEA Evaluations of Relative Efficiencies of Public and Private Institutions of Higher Learning," *Socio-Economic Planning Science* (forthcoming, 1989b).

Aigner, D.J., T. Amemiya and P.J. Poirier, "On the Estimation of Production Frontiers: Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function," *International Economic Review XVII* (1976) pp. 377–396.

Aigner, D.J. and S.F. Chu, "On Estimating the Industry Production Function," *American Economic Review LVIII* (1968) pp. 826–839.

Aigner, D.J., C.A. Lovell and P. Schmidt, "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics VI* (1977) pp. 21–37.

Ali, I., A. Charnes, W.W. Cooper, D. Divine, G.A. Klopp and J. Stutz, "An Application of Data Envelopment Analysis to Management of Army Recruiting Districts," Research Report CCS 436 (Austin, Texas: The University of Texas at Austin, Center for Cybernetic Studies, 1985).

Anselin, L., and J.S. Henderson, *A Decision Support System for Utility Performance Evaluation* (Columbus, Ohio: The National Regulatory Research Institute, 1985).

Banker, R.D., "A Game Theoretic Approach to Measuring Efficiency," *European Journal of Operational Research* 5 (1980a) pp. 262–270.

Banker, R.D., "Estimating Most Productive Scale Size Using Data Envelopment Analysis," *European Journal of Operational Research* 217 (1984) pp. 35–40.

Banker, R.D., "Models for the Measurement of Quality, Price and Cost Efficiencies," Working Paper (Carnegie Mellon University, School of Urban and Public Affairs, 1985).

Banker, R.D., "Productivity and Profitability Ratio Analysis," *The Accounting Review* (submitted 1988).

Banker, R.D., "Studies in Cost Allocation and Efficiency Evaluation," DBA Thesis (Harvard University Graduate School of Business, 1980b). Available from University Microfilms International, Ann Arbor, Michigan.

Banker, R.D., A. Charnes and W.W. Cooper, "Models for Estimating Technical and Scale Efficiencies," *Management Science* 30, 1984, pp. 1078–1092.

Banker, R.D., A. Charnes, W.W. Cooper and A. Maindiratta, "A Comparison of DEA and Translog Estimates of Production Frontiers Using Simulated Observations from a Known Technology," in A. Dogramaci and R. Färe, eds., *Applications of Modern Production Theory: Efficiency and Productivity* (Norwell MA: Klüwer Academic Publishers, 1987).

Banker, R.D., R.F. Conrad and R.P. Strauss, "A Comparative Application of DEA and Translog Methods: An Illustrative Study of Hospital Production," *Management Science* 36 (1986), pp. 30–34.

Banker, R.D. and A. Maindiratta, "Nonparametric Analysis of Technical and Allocative Efficiencies in Production," *Econometrica* 56 (1988), pp. 1315–1332.

Banker, R.D. and A. Maindiratta, "Piecewise Loglinear Estimation of Efficient Production Surfaces," *Management Science* 32 (1986) pp. 126–135.

Banker, R.D. and R.C. Morey, "Efficiency Analysis for Exogenously Fixed Inputs and Outputs," *Operations Research* 34 (1986) pp. 513–521.

Banker, R.D. and R.M. Thrall, "Estimating Returns to Scale in Data Envelopment Analysis," Working Paper (Carnegie Mellon University School of Urban and Public Affairs, 1988).

Bitran, G. and J. Valor-Sasbatier, "Some Mathematical Programming Based Measures of Efficiency in Health Care Institutions," in J.B. Guerard, Jr., G.D. Reeves and K. Laurence, eds., *Advances in Mathematical Programming and Financial Planning* (Greenwich CT: JAI Press, 1987).

Bowlin, W.F., "A Data Envelopment Analysis Approach to Performance Evaluation in Not-for-Profit Entities with an Illustrative Application to Base Maintenance Activities in the U.S. Air

Force," Ph.D. Thesis (The University of Texas at Austin, TX, Graduate School of Business, 1984). Available from University Microfilms International, Ann Arbor, Michigan.

Bowlin, W.F., J. Brennan, A. Charnes, W.W. Cooper and T. Sueyoshi, "A Model for Measuring Amounts of Efficiency Dominance," Research Report Center for Cybernetic Studies, Austin, Texas (The University of Texas at Austin, 1984).

Charnes, A., C.T. Clark, W.W. Cooper and B. Golany, "A Developmental Study of Data Envelopment Analysis in Measuring the Efficiency of Maintenance Units in the U.S. Air Force," *Annals of Operations Research* (1985) pp. 95–112.

Charnes, A. and W.W. Cooper, "Management Models and Industrial Applications of Linear Programming," *Management Science* 4 (1957) pp. 38–91.

Charnes, A. and W.W. Cooper, *Management Models and Industrial Applications of Linear Programming* (New York: John Wiley and Sons, 1961).

Charnes, A. and W.W. Cooper, "Preface to Topics in Data Envelopment Analysis," in R. Thompson and R. Thrall, eds., *Annals of Operations Research* 2 (1985) pp. 59–94.

Charnes, A. and W.W. Cooper, "Programming with Linear Fractional Functionals," *Naval Research Logistics Quarterly* 9 (1962) pp. 181–186.

Charnes, A., W.W. Cooper, D. Divine, G. Klopp and J. Stutz, "An Application of Data Envelopment Analysis to U.S. Army Recruitment Districts," Research Report (The University of Texas at Austin, Center for Cybernetic Studies, 1982).

Charnes, A., W.W. Cooper and B. Golany, "Relative Effects by Data Envelopment Analysis of Service Specific and Joint National Advertising in Navy Recruitment Activities," Research Report (The University of Texas at Austin, Center for Cybernetic Studies, 1986a).

Charnes, A., W.W. Cooper and B. Golany, "Relative Effects of Service Specific and Joint National Advertising in Marine Corps Recruitment Activities," Research Report (The University of Texas at Austin, Center for Cybernetic Studies, 1986b).

Charnes, A., W.W. Cooper, B. Golany, R. Halek, G. Klopp, E. Schmitz and D. Thomas, "Data Envelopment Analysis Approaches to Policy Evaluation and Management of Army Recruiting Activities I: Tradeoffs Between Joint Services and Army Advertising," Research Report CCS 532 (The University of Texas at Austin, Center for Cybernetic Studies, 1986c).

Charnes, A., W.W. Cooper, B. Golany, L. Seiford and J. Stutz, "Foundations of Data Envelopment Analysis and Pareto-Koopmans Efficient Empirical Production Functions," *Journal of Econometrics* 30 (1985) pp. 91–107.

Charnes, A., W.W. Cooper, Z.M. Huang and D.B. Sun, "Polyhedral Cone-Ratio DEA Models and Managerial Performance of Large Commercial Banks," in A. Lewin and K. Lovell, eds., *Proceedings of a Conference on Frontier Estimation* (forthcoming 1989).

Charnes, A., W.W. Cooper, D.B. Learner and F.Y. Phillips, "Management Science and Marketing Management," *Journal of Marketing* 49 (1985) pp. 93–105.

Charnes, A., W.W. Cooper and E. Rhodes, "Management Science Relations for Evaluation and Management Accountability," *Journal of Enterprise Management* 2 (1980) pp. 143–162.

Charnes, A., W.W. Cooper and E. Rhodes, "Measuring Efficiency of Decision Making Units," *European Journal of Operational Research* 1 (1978) pp. 429–444.

Charnes, A., W.W. Cooper, J. Rousseau and J. Semple, "Data Envelopment Analysis and Axiomatic Notions of Efficiency and Reference Sets," Research Report CCS 558 (The University of Texas at Austin Center for Cybernetic Studies, 1987).

Charnes, A., W.W. Cooper, L. Seiford and J. Stutz, "Invariant Multiplicative Efficiency and Piecewise Cobb-Douglas Envelopments," *Operations Research Letters* 2 (1983) pp. 101–105.

Charnes, A., W.W. Cooper and T. Sueyoshi, "A Goal Programming/Constrained Regression Review of the Bell System Breakup," *Management Science* 34 (1988) pp. 1–38.

Charnes, A., W.W. Cooper and R.M. Thrall, "Identifying and Classifying Scale and Technical Efficiencies and Inefficiencies in Observed Data Via Data Envelopment Analysis," *Operations Research Letters* 5 (1986) pp. 105–110.

Charnes, A., W.W. Cooper and R.M. Thrall, "A Structure for Characterizing and Classifying Efficiencies and Inefficiencies in Data Envelopment Analysis," *Journal of Productivity Analysis* (submitted, 1988).

Charnes, A. and S. Zlobec, "Stability of Efficiency Evaluations in Data Envelopment Analysis," *Zeitschrift für Operations Research*, Series A: Theory (forthcoming, 1989).

Clarke, R., "Effects of Repeated Applications of Data Envelopment Analysis on Efficiency of Air Force Vehicle Maintenance Units in the Tactical Air Command and a Test for the Presence of Organizational Slack Using Rajiv Banker's Game Theory Formulations," Ph.D. Thesis, (The University of Texas at Austin Graduate School of Business, 1988).

Cooper, W.W. and Y. Ijiri, eds., *Kohler's Dictionary for Accountants, 6th edition* (Englewood Cliffs, NJ: Prentice-Hall, Inc., 1983).

Dewald, W.G., J.G. Thursby and R.G. Anderson, "Replication of Empirical Economics," *The American Economic Review* 76 (1986) pp. 587–603.

Färe, R., *Fundamentals of Production Theory* (Berlin: Springer-Verlag, 1988).

Färe, R., S. Grosskopf and C.A. Knox Lovell, *The Measurement of Efficiency of Production* (Boston: Kluwer-Nijhoff, 1985).

Farrell, M.J., "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, Series A (1957) pp. 253–290.

Farrell, M.J. and M. Fieldhouse, "Estimating Efficient Production Frontiers Under Increasing Returns to Scale," *Journal of the Royal Statistical Society* Series A (1962) pp. 252–267.

Lalonde, R.J., "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *The American Economic Review* 76 (1986) pp. 604–619.

Leibenstein, H., *Beyond Economic Man* (Cambridge: Harvard University Press, 1976).

Lovell, C.A.K. and R. Sickles, "Testing Efficiency in Joint Production: A Parametric Approach," *Review of Economics and Statistics* 65 (1983) pp. 51–58.

Rhodes, E., "Data Envelopment Analysis and Related Approaches for Measuring Efficiency of Decision Making Units with an Application to Program Follow Through in U.S. Education," Ph.D. Thesis (Carnegie Mellon University, School of Urban and Public Affairs, 1978). Available from University Microfilms International, Ann Arbor, Michigan.

Rhodes, E. and L. Southwick, "Determinants of Efficiency in Public and Private Universities," Mimeo Research Report (University of Indiana School of Environmental and Public Affairs, 1986).

Sager, T., "Estimating Modes and Isopleths," *Communications in Statistics-Theory and Methods* 12, No. 5 (1983) pp. 529–557.

Sager, T. and Thisted, R.A., "Maximum Likelihood Estimation of Isotonic Modal Regression," *The Annals of Statistics* 10, No. 3 (1982) pp. 690–707.

Schaible, S., "Parameter-free Convex Equivalent and Dual Programs of Fractional Programming Problems," *Zeitschrift für Operations Research* 18 (1974) pp. 187–196.

Shephard, R.W., *The Theory of Cost and Production Functions* (Princeton: Princeton University Press, 1970).

Sherman, D., "Measurement of Hospital Technical Efficiency: A Comparative Evaluation of Data Envelopment Analysis and Other Efficiency Measurement Techniques," DBA Thesis (Harvard University, Graduate School of Business, 1980). Available from University Microfilms International, Ann Arbor, Michigan.

Stigler, G.J., "The Xistence of X-Efficiency", *American Economic Review* LXVI (1976) pp. 213–216.

Sudit, E.F., *Productivity Based Management* (Boston: Kluwer-Nijhoff Publishing Co., 1984).

Thomas, D., *Auditing the Efficiency of Regulated Companies: An Application of DEA to Electric Cooperatives* (IC[2] Institute, The University of Texas at Austin, 2185 San Gabriel St., Austin, TX 78705, 1986).

Thompson, R.G., F.D. Singleton, Jr., R.M. Thrall and B.A. Smith, eds., "Comparative Site Evaluations for Locating a High Energy Physics Lab [i.e., the Supercollider] in Texas," *Interfaces* 16 (1986) pp. 35–49.

Thrall, R.M., "Overview and Recent Developments in DEA: The Mathematical Programming Approach," in A. Lewin and K. Lovell, eds., *Proceedings of a Conference on Frontier Estimation* (forthcoming, 1989).