MANAGEMENT SCIENCE
Vol. 27, No. 12, December 1981
Printed in U.S.A

A BI-EXTREMAL PRINCIPLE FOR FRONTIER ESTIMATION AND EFFICIENCY EVALUATIONS*

R. D. BANKER,† A. CHARNES,‡, W. W. COOPER‡ AND A. P. SCHINNAR§

A new approach is supplied for locating efficiency frontiers and evaluating the efficiency of Decision Making Units (DMU's). This is accomplished from observational data by means of an envelopment procedure called DEA (Data Envelopment Analysis) originally developed by Charnes, Cooper and Rhodes [10] in connection with their ratio formulation for relative efficiency measurement. The current variant employs a bi-extremal principle which, though nonlinear, is subsequently shown to be reducible to a finite sequence of linear programming problems. The development is illustrated by means of multiple output functions which are piecewise of Cobb-Douglas or general log linear type and which also allow for increasing, decreasing and constant returns to scale. More than one production function for the DMU's is also allowed. The reduction of the bi-extremal principle to linear programming equivalence is also accomplished for much more general classes of functions. A numerical example illustrates some of these ideas and also provides a springboard for a new theorem which relates these efficiency measures to ones which were supplied earlier in the Charnes, Cooper and Rhodes's work [10].

(EFFICIENCY; DECISION MAKING UNITS; PRODUCTION FUNCTIONS; EFFICIENCY FRONTIERS)

1. Introduction

Reference [10][1] provided a nonlinear ratio extremal principle for determining relative efficiency for a collection of DMU's (Decision Making Units) from observational data. The associated dual analysis determined immediately, via a general concept of Data Envelopment Analysis (DEA), a local approximation, piecewise linear to an assumed (but unknown) underlying production function in the neighborhood of efficient input-output DMU's. See [10]. These DEA procedures were also used to distinguish between "Program," "Managerial" and other types of efficiency. See [9]. Of course, the use of the "CCR ratio" does not imply that the underlying DMU production functions are linear and, indeed, these need not be specified in parametric form at all in a DEA approach. On the other hand, in some situations one may have reason to assume that parametric families of functions are involved which are assumed to represent the production functions in at least a piecewise manner. One may then wish to estimate the parameters by a DEA principle with assurance that the resulting functions give the correct efficiency measures over the relevant pieces of the efficiency frontier.

In this paper we shall develop a bi-extremal principle for achieving such parametric estimates and an associated efficiency measure to be assigned to each DMU. The basic idea is as follows: We "envelop" the observed values of the outputs by means of

*Accepted by Arie Y. Lewin; received April 23, 1980. This paper has been with the authors 7 months for 2 revisions.

†Carnegie-Mellon University.
‡University of Texas at Austin.
§University of Pennsylvania.

[1] The numbers in square brackets refer to the correspondingly numbered items in the bibliography.

0025-1909/81/2712/1370$01.25

parametric functions of the observed input values. The envelope is "tightened" to rest on observed output values by means of a minimization operation. Then the parametric output functions are used to replace the observed output values in the manner of the nonlinear ratio maximization principle given in [10].[2] One thus obtains a bi-extremal principle of "maximin" type for the local technology of each DMU. With this in hand, a determination of efficiency for each DMU is then made under an appropriate modification of the basic concepts in [10].

For output functions of Cobb-Douglas type (or, more generally, for "ϕ linear" type)[3] we show in this paper that this nonlinear-nonconvex bi-extremal principle is, in fact, equivalent to a finite sequence of linear programming problems with one nonlinear vector operation. The linear programs further differ only in the right-hand side, thus greatly facilitating computation and analysis.

Most current approaches to the estimation of extreme values (e.g., as in mathematical statistics) are restricted to zero dimensional values (or frontiers) such as the end points of a uniform distribution. Our bi-extremal principle provides a new approach to the determination of multi-dimensional (extreme) frontiers such as are required for efficiency determination in the multiple input-multiple output situations that are commonly encountered in public policy evaluation problems. The principle is more general than such multiple input-multiple output applications, however, and extends to any case where extreme frontiers are of interest.

2. Background

We follow the notation conventions of [10] and consider the common input and output values for a collection of DMU's defined as follows:

$$\begin{aligned} x_{ij} &= \text{the amount of input}, \quad i = 1, \ldots, m; \\ y_{rj} &= \text{the amount of output}, \quad r = 1, \ldots, s; \end{aligned} \tag{1}$$

where each $j = 1, \ldots, n$ indexes one of the DMU's being considered. These x_{ij} and y_{rj} values will generally represent observations generated from past behavior and we shall assume that they all have positive values.

The following formulation was given in [10] for determining the efficiency of any specified DMU_0 from among this set of $j = 1, \ldots, n$ DMU's:

$$\begin{aligned} \max h_0 &= \frac{\sum_{r=1}^{s} w_r y_{r0}}{\sum_{i=1}^{m} u_i x_{i0}} \\ \text{subject to} \quad & \frac{\sum_{r=1}^{s} w_r y_{rj}}{\sum_{i=1}^{m} u_i x_{ij}} \leq 1, \qquad j = 1, \ldots, n, \\ & w_r, u_i > 0, \quad \forall i, r. \end{aligned} \tag{2}$$

Because the ratio in the functional also appears in the constraints we have max. $h_0 = h_0^* \leq 1$ in any case and, as shown in [10], $h_0^* = 1$, if and only if DMU_0 is efficient. Note, in particular, that this provides a scalar value for the wanted efficiency rating.[4]

[2] See the expression in (2) below.
[3] See §7 below.
[4] Discussions of the operational significance of this rating, as well as a transformation into linear programming equivalents (e.g., for computational efficiency) are provided in [10].

This scalarization is achieved via the positive weights w_r, u_i assigned to the respective outputs and inputs. These weights are not assigned *a priori* in an arbitrary manner, however, but are determined objectively from the data as prescribed by (2). Hereafter we shall refer to them as "virtual weights"—the intended analogy being to concepts like "virtual displacements" and/or "virtual work" in physics or engineering,[5] which represent magnitudes that are not observed directly but are implicit, instead, in the underlying physical principles and models.

In the present paper we wish to consider situations in which we have *a priori* information on the "technological possibilities" of input to output transformations by each DMU. This information consists of knowledge of the functional form of the transformation for various subsets of the DMU's. We consider chiefly the case where the technology is such that the efficiency frontier function is piecewise log linear with emphasis on functions which are of Cobb-Douglas type—whose importance has been validated to an extraordinary degree in a large number of empirical studies in many different periods and in many different countries with different economic settings.[6] Related work on production function representations, in extended Cobb-Douglas format—[12], [13],— makes it clear that this extends to a very general class of cases. The fact that other functions, e.g., the translog and other flexible functional forms, require such piecewise representation in the presence of simple capacity constraints (see [14]) also makes it desirable to proceed as we propose to do.

We need to determine both the local pieces of the efficient function and to rate the efficiency of each DMU. To do the latter, we form the ratio of virtual outputs to virtual inputs as determined from observed data. This would be sufficient for an efficiency measure without recourse to further constraining relations if we were dealing with simple efficiency concepts, as in many parts of science and engineering. For instance, we would not need to explicitly impose conditions to insure that potential technological output cannot exceed input when dealing with electrical energy transformed into incandescent light energy since thermodynamical principles insure the presence of degradation. When dealing with the virtual output to input ratios for the complex (multiple) inputs and outputs involved in our DMU's, however, we need to ensure this "degradation" by requiring the ratios to be between zero and unity. Because all outputs and inputs are positive and because the "weights", to be determined from these observational data by our extremization principle are also all positive, the ratios are automatically positive. The efficiency values we obtain—which are then only relative efficiencies for each DMU—will have by the degradation conditions a limit of unity in all cases.

To estimate the output function for an individual DMU, we employ a bi-extremal "maximin principle" which simultaneously takes account of our information on all of the other DMU's. These, as we have already noted, are assumed to have a "productive technology" of the same functional form. We posit an "envelopment condition" on the observed outputs of all DMU's such that the observed inputs are transformed by this function into outputs which are no less than the corresponding observed values. See (3), below. We tighten this envelope by means of a minimization operation on the

[5]See also the discussion on p. 647 in [6]. Further discussion of these points is continued at the end of §3 where they can be further clarified by reference to the intervening mathematical development.

[6]See Walters [17, p. 328].

functional with respect to the parameters of this functional form. The determination of the parameter values is completed by maximinizing a ratio of virtual output (in functional form) to virtual input. Thus we have our "maximin" or bi-extremal principle for determining the output function.

The sense of what has just been said can be clarified, at least in a general way, by comparing (4), below, with (2), above. The ratio in the latter, i.e., (2), involves directly observed outputs and inputs. In the case of (4) we continue with the directly observed inputs in the denominator, as before, but replace in the functional the directly observed outputs in the numerator of (2) by the posited functional form with parameters A_r and μ_{ri}.

The minimization in the maximin operator is, as already noted, with respect to the parameters of the posited functional form which, because of the form, insures that they exceed the observed outputs by as little as possible. This is followed by maximization with respect to the "virtual weights" w_r and u_i which are the components (all positive) of the vectors w and u. The constraints consist of (a) the "envelopment conditions" which are made tight by the minimization operation (b) the degradation conditions on the observed outputs and inputs and (c) other relevant conditions on the parameters of the function and the virtual weights.[7]

Continuing our comparison between (2) and (4), we observe that the former gives the wanted efficiency measure directly. This is not the case for (4). The maximin value obtained for the particular DMU being rated by this bi-extremal principle is not an efficiency measure by virtue of the following two considerations: (a) the ratio in (4) is not one of virtual observed outputs to virtual observed inputs, as in (2), and (b) the degradation condition in the constraints do not insure that the virtual potential (technologically possible) output does *not* exceed the virtual input for *all* DMU's.

The ways in which such an efficiency measure can be achieved will now be undertaken along with related ideas (and extensions) in the mathematical development that follows.

3. Development

To help fix the ideas, we present our developments first in terms of Cobb-Douglas types of functions, and then present the more general formulae later in the paper. Thus, we now introduce the following "envelopment condition" on the outputs,

$$y_{rj} \leqslant A_r \prod_{i=1}^{m} x_{ij}^{\mu_{ri}}, \qquad j = 1, \ldots, n, \tag{3}$$

where the variables μ_{ri} and A_r are constrained to be non-negative.[8] Here the symbol $\prod_{i=1}^{m}$ refers to the m-fold product of the $x_{ij}^{\mu_{ri}}$ values so that although we are now restricting attention to functions which are of Cobb-Douglas form, we are also extending them to the case of multiple outputs—$r = 1, \ldots, s$ in number.[9]

[7] Usually in the form of positivity requirements including, possibly, lower and upper bounds on these positive values.

[8] Recall that y_{rj} and x_{ij} are observed positive constants.

[9] See [12] and [13] for further extension to the class of positive homogeneous and analytic functions for these Cobb-Douglas type of formulations.

Our bi-extremal principle for determination of the output function for any particular DMU is

$$\max_{w,u} \min_{A_r, \mu_{ri}} \frac{\sum_{r=1}^{s} w_r A_r \prod_{i=1}^{m} x_{i0}^{\mu_{ri}}}{\sum_{i=1}^{m} u_i x_{i0}}$$
$$\text{subject to} \quad y_{rj} \leqslant A_r \prod_{i=1}^{m} x_{ij}^{\mu_{ri}}, \qquad r = 1, \ldots, s$$
$$\sum_{r=1}^{s} w_r y_{rj} \leqslant \sum_{i=1}^{m} u_i x_{ij}, \qquad j = 1, \ldots, n \tag{4}$$
$$A_r, \mu_{ri} \geqslant 0; \; w_r, u_i > 0, \qquad \forall r, i;$$

where the DMU designated by "0" is represented in the functional as well as the constraints.[10]

Applying this determination to each DMU, we obtain the n output-input functions

$$\eta_r^j(x) = A_r^*(j) \prod_{i=1}^{m} x_i^{\mu^*_{ri}(j)}, \qquad j = 1, \ldots, n, \tag{5}$$

where $x = (x_1, \ldots, x_m)$ represents possible input values, and the $A_r^*(j)$ and $\mu_{ri}^*(j)$ denote the $r = 1, \ldots, s$ parameter values obtained from the maximin procedures.

Because the multiple outputs were simultaneously determined from the same inputs, it follows that, in principle, the $r = 1, \ldots, s$ relations specified in (5) should also be simultaneously estimated from these same data. This principle is incorporated in (4) although, as will be shown, we can provide a development which makes it possible to use a series of ordinary linear programming problems in which each of these $r = 1, \ldots, s$ relations may also be estimated separately.

As was true in (2), the efficiency evaluations for each DMU are determined relative to the data for all of the DMU's. The original y_{rj} data in (2) are now, however, replaced by the estimates in (5) for each efficiency evaluation via

$$\max_{w,u} \frac{\sum_{r=1}^{s} w_r y_{r0}}{\sum_{i=1}^{m} u_i x_{i0}}$$
$$\text{subject to} \quad \sum_{r=1}^{s} w_r A_r^*(j) \prod_{i=1}^{m} x_{ij}^{\mu^*_{ri}(j)} \leqslant \sum_{i=1}^{m} u_i x_{ij}, \qquad j = 1, \ldots, n; \tag{6}$$
$$w_r, u_i > 0, \qquad \forall r, i.$$

In other words, the efficiency evaluation for each DMU is now effected relative to the optimal output-input relations that were estimated to obtain for every DMU.

We can now observe that we have preserved the output-to-input ratio form, as in (2), with the constraints ensuring that the resulting (scalar) efficiency rating will not exceed unity. Our present formulations thus retain contact with the developments in [10] which were shown to generalize the usual science and engineering definitions of efficiency while also extending them to situations involving multiple outputs and inputs.

[10] See the "Short Communication" referenced in [10] for a discussion of the reasons for restricting the w_r and u_i to positive values, and see [8] for a development in terms of a "Non-Archimedean Model Theorem."

The efficiency rating for each DMU evidently depends on the choice of the optimal w_r and u_i values in (6). For the applications associated with (2), it was shown in [10][11] that these variables play a central role not only in (i) determining the efficiency rating of the DMU being evaluated but also in (ii) relating these evaluations to generalizations of basic concepts like efficient productivities, etc., in the economic theory of production as well as in (iii) providing a new way for estimating the parameters of such extremal relations from the dual pair of ordinary linear programming problems that can be associated with (2) via the theory of fractional programming.

We shall follow a similar route here and also utilize the theory of fractional programming to develop a *series* of linear programming problems to effect the efficiency evaluations that are wanted in (6). At this juncture, however, it may be well to emphasize that these w_r and u_i values are not "weights" which are assumed to apply via some *a priori* assumption of relative importance or via some imputation from extraneous data. They are, instead, to be objectively determined from the data of the problem in terms of the explicitly stated models, optimizing objectives and constraints. We, therefore, underscore this, as in [6] and [3], by referring to these w_r and u_i values as "virtual transforms" or "virtual rates of transformation" which when applied to the original data (as in the ratio form of the functional in (6)) result in a corresponding "virtual output" to "virtual input" ratio. At an optimum the resulting w_r^* and u_i^* values are then distinguished as the "efficient transforms" with the resulting virtual output to virtual input ratio then providing the efficiency evaluation that is wanted.

All of this is only a formality, however, unless we can do something about the formulation in (4) which involves non-convexity (or non-concavity) as well as bi-extremality in the objective. We therefore now proceed to show how to reduce the solution of (4) to an *a priori* fixed finite sequence of linear programming problems with one nonlinear vector operation.

4. Transformation

We introduce new variables z_{r0} via

$$A_r \prod_{i=1}^{m} x_{i0}^{\mu_{ri}} \leq z_{r0}, \qquad r = 1, \ldots, s. \tag{7}$$

As a result of adjoining these additional constraints to (4) an equivalent objective function is

$$\max_{w,u} \min_{z_{r0}} \frac{\sum_{r=1}^{s} w_r z_{r0}}{\sum_{i=1}^{m} u_i x_{i0}}. \tag{8}$$

Next we make use of the following change of variables,

$$\begin{aligned} w_r' &= t w_r, \\ u_i' &= t u_i, \\ t &\geq 0 \end{aligned} \tag{9}$$

such that

$$\sum_{i=1}^{m} u_i' x_{i0} = 1. \tag{10}$$

[11] See also [6] and [4] and [37].

Multiplying the numerator and denominator of (8) by t and doing the same in the constraints involving the w_r and u_i values in (4), we can also utilize the relations (7) to write

$$\max_{w', u'} \min_{z_{r0}} \sum_{r=1}^{s} w'_r z_{r0}$$

$$\text{subject to} \quad A_r \prod_{i=1}^{m} x_{i0}^{\mu_{ri}} \leq z_{r0},$$

$$y_{rj} \leq A_r \prod_{i=1}^{m} x_{ij}^{\mu_{ri}},$$

$$\sum_{r=1}^{s} w'_r y_{rj} \leq \sum_{i=1}^{m} u'_i x_{ij}, \tag{11}$$

$$\sum_{i=1}^{m} u'_i x_{i0} = 1,$$

$$A_r, \mu_{ri} \geqslant = 0; w'_r, u'_i > 0; r = 1, \ldots, s; i = 1, \ldots, m; j = 1, \ldots, n.$$

Via the theory of fractional programming, as in [7], we can evidently use the relations (9) and (10) to move from (11) to (4), as required, and so we can now continue our development from (11) as follows.

By taking logs in the first two inequalities, we can rewrite (11) in the form:

$$\max_{w', u'} \min_{z_{r0}} \sum_{r=1}^{s} w'_r z_{r0}$$

$$\text{subject to} \quad -\ln(z_{r0}) + \hat{A}_r + \sum_{i=1}^{m} \mu_{ri} \hat{x}_{i0} \leqslant 0,$$

$$-\hat{A}_r - \sum_{i=1}^{m} \mu_{ri} \hat{x}_{ij} \leqslant -\hat{y}_{rj}, j = 1, \ldots, n, \tag{12}$$

$$\sum_{i=1}^{m} u'_i x_{ij} = 1,$$

$$\sum_{r=1}^{s} w'_r y'_{rj} - \sum_{i=1}^{m} u'_i x_{ij} \leqslant 0,$$

$$\mu_{ri} \geqslant 0; w'_r, u'_i > 0; r = 1, \ldots, s; i = 1, \ldots, m;$$

where the caret above a symbol denotes natural logarithm.

5. Linear Programming Equivalent

Notice in (12) that the w', u' and the $z_{r0}, \hat{A}_r, \mu_{ri}$ sets of variables occur in separate, independent sets of constraints. These two constraint systems restrict their separate variables to different feasible sets. Further, with the bilinear functional, the bi-extremal problem is equivalent to a saddle-value problem and such an equivalence would hold for much more general bi-extremal problems than (12). We shall pursue this generality in subsequent papers.

Here we note next that the $z_{r0}, \hat{A}_r, \mu_{ri}$ constraint system separates into independent systems with one system for each output index r. Thus, for any set of the nonnegative

w'_r values, we wish to minimize *each* z_{r0} subject to its separate rth constraint system. Setting $\hat{z}_{r0} = \ln(z_{r0})$, the rth problem would be:

$$\begin{aligned} \min \quad & e^{\hat{z}_{r0}} \\ \text{subject to} \quad & \hat{z}_{r0} - \hat{A}_r - \sum_{i=1}^{m} \mu_{ri}\hat{x}_{i0} \geq 0, \\ & \hat{A}_r + \sum_{i=1}^{m} \mu_{ri}\hat{x}_{ij} \geq \hat{y}_{rj}, \quad j = 1, \ldots, n, \\ & \mu_{ri} \geq 0, \quad i = 1, \ldots, m. \end{aligned} \tag{13.1}$$

It should be recalled that we are here explicitly writing down only the generally essential constraints. Typically, further constraints on the A_r and the μ_{ri} would be imposed such as $\hat{A}_r < K_r$ and $\mu_{ri} \leq U$. In any case, we can proceed to obtain a linear programming problem from (13.1) by observing that minimizing $e^{\hat{z}_{r0}}$ is equivalent to minimizing $\hat{z}_{r0}$. Thus, we have the linear programming problem:

$$\begin{aligned} \min \quad & \hat{z}_{r0} \\ \text{subject to} \quad & \hat{z}_{r0} - \hat{A}_r - \sum_{i=1}^{m} \mu_{ri}\hat{x}_{i0} \geq 0, \\ & \hat{A}_r + \sum_{i=1}^{m} \mu_{ri}\hat{x}_{ij} \geq \hat{y}_{rj}, \quad j = 1, \ldots, n, \\ & \mu_{ri} \geq 0, \quad i = 1, \ldots, m. \end{aligned} \tag{13.2}$$

Hence, as promised, we have specified an *a priori* sequence of linear programming problems together with one nonlinear operation (converting $\hat{z}^*_{r0}$ to z^*_{r0}) to solve our combined problem of determining a piecewise Cobb-Douglas type efficiency function. This is then used for the efficiency ratings of each DMU as described in the paragraph leading into (14) below.

Notice that the functional and matrix coefficient structure for this linear programming system is the same for all $r = 1, \ldots, s$. Only the right hand side, the $\hat{y}_{rj}$, would change with r. If one were to solve the system via the dual linear programming problems, only the functional would change with r. One then knows *a priori* each function to be employed. Thus an optimal basis for one value of r would be a feasible basis for any value of r.

On obtaining the optimal $\hat{z}_{r0} = \hat{z}^*_{r0}$, for each $j = 1, \ldots, n$, we make the inverse (exponential) transformation to get z^*_{r0} which, of course, are the $A^*_{r(0)} \prod_{i=1}^{m} x_{i0}^{\mu^*_{ri}(0)}$. Our efficiency determination problems are then

$$\begin{aligned} & \text{maximize } h_0 = \sum_{r=1}^{s} w'_r y_{r0} \\ & \text{subject to} \quad \sum_{r=1}^{s} w'_r z^*_{rj} - \sum_{i=1}^{m} u'_i x_{ij} \leq 0, \quad j = 1, \ldots, n, \\ & \qquad \sum_{i=1}^{m} u'_i x_{i0} = 1, \\ & \qquad w'_r, u'_i, > 0; \quad \forall r, i; \end{aligned} \tag{14}$$

which can evidently be related to the development in [10] by reference to the transformation of (2) to a linear programming equivalent and replacement of y_{rj} by z_{rj}^*.

6. Numerical Illustration

Preliminary to extending these ideas we stop to illustrate these developments by reference to five hypothetical DMU's with the observed values of their two inputs and two outputs that are arrayed in the following table:

TABLE 1

	DMU	1	2	3	4	5
Inputs	x_1	3	5	1	3	3
	x_2	3	1	5	3	3
Outputs	y_1	2	5	1	5	1
	y_2	2	1	5	1	5

We shall restrict our illustration to evaluating the efficiency of DMU_1, which we do by inserting the data from the above table into (13.2) and obtain

$$\begin{aligned} \min \quad & \hat{z}_{10} \\ \text{subject to} \quad & \hat{z}_{10} - \hat{A}_1 - \mu_{11} \ln 3 - \mu_{12} \ln 3 \geqslant 0 \\ & \hat{A}_1 + \mu_{11} \ln 3 + \mu_{12} \ln 3 \geqslant \ln 2 \\ & \hat{A}_1 + \mu_{11} \ln 5 + \mu_{12} \ln 1 \geqslant \ln 5 \\ & \hat{A}_1 + \mu_{11} \ln 1 + \mu_{12} \ln 5 \geqslant \ln 1 \\ & \hat{A}_1 + \mu_{11} \ln 3 + \mu_{12} \ln 3 \geqslant \ln 5 \\ & \hat{A}_1 + \mu_{11} \ln 3 + \mu_{12} \ln 3 \geqslant \ln 1 \\ & \mu_{11}, \mu_{12} \geqslant 0. \end{aligned} \tag{15}$$

where, as in the development from (7) to (13.2) the $\hat{A}_1 \equiv \ln A$, and μ_{11}, μ_{12} values represent the parameters of the production function to be estimated for the corresponding level of the first output that should have been generated by DMU_1 if it were being managed efficiently. This output level is obtained from $\hat{z}_{10} = \ln z_{10}$ with, of course, an optimal $z_{10} = z_{10}^* \geqslant y_1 = 2$ the output that was actually attained by DMU_1.

Evidently the data from all DMU's enter into this efficiency evaluation with, nevertheless, this one piece of the production surface being shaped particularly for the evaluation of DMU_1. That is, we are using the entire set of available data but in a way that differs from the usual industry type of production function estimation approaches in economics (or other disciplines) and it also differs from the one-firm-at-a-time approaches that have also been commonly employed. In contrast to the latter approach, we are not restricting ourselves only to the firm's own record of accomplishment but are, instead, employing data from all firms to obtain a function that will apply locally and make it possible to evaluate those accomplishments. In contrast to the former approach (i.e., via industry-wide functions) we are shaping each piece of the function for evaluation of the decision making units, as pertinent, and not searching for one global function that has no decision making unit associated with it.

The determination of these z^*_{r0} values is, of course, only a first stage for determining our wanted efficiency evaluations. To show what else is involved we first rewrite (15) in the form

$$\begin{aligned} \min \quad & \hat{z}_{10} = \hat{A}_1 + \mu_{11}\ln 3 + \mu_{12}\ln 3 \\ \text{subject to} \quad & \hat{A}_1 + \mu_{11}\ln 3 + \mu_{12}\ln 3 \geqslant \ln 2 \\ & \hat{A}_1 + \mu_{11}\ln 5 + \mu_{12}\ln 1 \geqslant \ln 5 \\ & \hat{A}_1 + \mu_{11}\ln 1 + \mu_{12}\ln 5 \geqslant \ln 1 \\ & \hat{A}_1 + \mu_{11}\ln 3 + \mu_{12}\ln 3 \geqslant \ln 5 \\ & \hat{A}_1 + \mu_{11}\ln 3 + \mu_{12}\ln 3 \geqslant \ln 1 \\ & \mu_{11}, \mu_{12} \geqslant 0, \end{aligned} \tag{16.1}$$

with the associated dual linear program

$$\begin{aligned} \max \quad & w_1\ln 2 + w_2\ln 5 + w_3\ln 1 + w_4\ln 5 + w_5\ln 1 \\ \text{subject to} \quad & w_1 + w_2 + w_3 + w_4 + w_5 = 1 \\ & w_1\ln 3 + w_2\ln 5 + w_3\ln 1 + w_4\ln 3 + w_5\ln 3 \leqslant \ln 3 \\ & w_1\ln 3 + w_2\ln 1 + w_3\ln 5 + w_4\ln 3 + w_5\ln 3 \leqslant \ln 3 \\ & w_1, w_2, w_3, w_4, w_5 \geqslant 0. \end{aligned} \tag{16.2}$$

Proceeding via the latter problem we utilize the strong duality relations of ordinary linear programming[12] to obtain $\hat{z}^*_{10} = \ln z^*_{10} = \ln z^*_{10} = \ln 5$ and hence $z^*_{10} = 5$. Indeed, taking advantage of the remarks in the second paragraph following (13.2) we readily obtain the z^*_{ij} which are recorded in Table 2.

TABLE 2

	DMU	1	2	3	4	5
Maximal Outputs	z_1^*	5	5	1	5	5
	z_2^*	5	1	5	5	5

A straightforward application of (14) then yields $h_0^* = 2/5$ for DMU_1 which suffices to characterize it as inefficient. It is of interest to note also, however, that DMU_1 would also have been characterized as inefficient but with a value of $h_0^* = 2/3$ if we had applied the model (2) to the data of Table 1. This, in turn, raises a question of the relation between these two approaches which we formalize via the following:[13]

THEOREM. *The efficiency rating for any* DMU_0 *obtained from* (14) *cannot exceed the efficiency rating from* (2) *when both are applied to the same data.*

[12] For simplicity we are ignoring the Nonarchimedean terms that would be needed to guarantee the positivity for the w_i^* in this and the following discussion. See [8].

[13] We are indebted to a referee for suggesting that this be included here. An even more general result, however, is established in Chapter 6 of [1] where it is shown that the production possibility set associated with (2) requires only minimal assumptions and hence will generally yield higher efficiency ratings than other alternatives.

PROOF. This is obtained by applying the theory of fractional programming to (2) via the change of variables in (9) and (10) to secure

$$\begin{aligned} \max h_0 = & \sum_{r=1}^{s} w_r' y_{r0} \\ \text{subject to: } & \sum_{r=1}^{s} w_r' y_{rj} - \sum_{i=1}^{m} u_i' x_{ij} \leqslant 0 \\ & \sum_{i=1}^{m} u_i' x_{i0} = 1 \\ & w_i', u_i \quad > 0 \qquad \forall r, i, \end{aligned} \tag{17}$$

This is the same functional as in (14) and the two constraint sets differ only in the replacement of y_{rj} by z_{ij}^*. However, we always have $z^*_{rj} \geqslant y_{rj}$ for every r and j. Thereby, every set of w_r satisfying the constraints of (14) also satisfies the constraints of (17). Hence, $h_0^* = \sum_{r=1}^{s} w_r'^* y_{r0}$ in (14) cannot exceed the $h_0^* = \sum_{r=1}^{s} w_r'^* y_{r0}$ in (17). Q.E.D.

The ratios y_{rj}/z_{rj}^* might themselves be used as measures of efficiency that would be employed in circumstances such as when one wanted to focus on the efficiency attained for particular outputs by particular DMU's.[14] The concept of a "technology" represents a "general state of knowledge"[15], however, that is free of any particular DMU. Hence the representation in (14) is taken over all of the $j = 1, \ldots, n$ DMU's—as well as over all of the pertinent inputs and outputs—in a manner that provides the efficiency measure by reference to the DMU's which the evidence reveals are most efficient. Unless we have some way of securing information on this general state of knowledge which we can incorporate in the reference set[16], however, our optimizations will generally produce only a measure of relative efficiency—as we have already noted.

This returns us to the points of contact with our earlier development along with the concepts of "virtual outputs" and "virtual inputs" from which the efficiency measure emerges for each DMU, and this, in turn, relates the developments in this paper to earlier work such as is reported in [10] and [15]. Thus, with this contact effected we can now turn to possible further generalizations and extensions in the sections that follow.

7. Generalization

It is clear that much more general output functions than those of Cobb-Douglas or general log linear type can be used to get equivalent linear programming systems of the form (13) and (14) by following the theory we have now provided. We develop these possibilities in the following manner. We shall call an output function $g(\theta_1, \ldots, \theta_{m_1}, x_1, \ldots, x_{m_2})$ "ϕ-linear" if:

$$\phi[g(\theta, x)] = \sum_{k=1}^{m_1} \theta_k f_k(x) \tag{18}$$

where ϕ is a monotone strictly increasing function.

[14] See the discussion in [3] and [6].
[15] See, e.g., H. A. Simon in [16].
[16] See the discussion in [10].

We shall assume each output function g_r is ϕ_r-linear. Applying the same techniques as in our Cobb-Douglas cases, we can evidently reduce our bi-extremal principle to solution of the two linear programming problems represented in (19.1) and (19.2) respectively. i.e.,

$$\begin{aligned} &\text{minimize} \quad \xi_{r0} = \phi_r\left[g_r(\theta, x^0) \right] \equiv \sum_{k=1}^{m_1} \theta_{rk} f_{rk}(x^0) \\ &\text{subject to} \quad \sum_{k=1}^{m_1} \theta_{rk} f_{rk}(x^j) \geqslant \phi_r(y^j), \qquad j = 1, \ldots, n \end{aligned} \tag{19.1}$$

where $\phi_r[g_r(\theta, x^j)] = \sum_{k=1}^{m_1} \theta_{rk} f_{rk}(x^j)$. Other relevant linear inequalities which do not involve other values of r may also be placed on the θ_{rk}.

With min $\xi_{r0} = \xi_{r0}^*$ available for all DMU's, i.e., ξ_{rj}^* for $j = 1, \ldots, n$, the efficiency system then becomes

$$\begin{aligned} &\max \quad \sum_{r=1}^{s} w_r' y_{r0} \\ &\text{subject to} \quad \sum_{i=1}^{m_2} u_i' x_{i0} = 1 \\ &\sum_{r=1}^{s} w_r' \phi_r^{-1}(\xi_{rj}^*) - \sum_{i=1}^{m_2} u_i' x_{ij} \leqslant 0 \\ &w_r', u_i' > 0. \end{aligned} \tag{19.2}$$

Thus, as was true for the Cobb-Douglas and log-linear cases, our two-stage procedure enables us to determine the parameters of the posited functional forms and also the efficiencies by reference to observed data.

Conclusion

The case of ϕ-linear functions includes, of course, both log linear and Cobb-Douglas families of functions. Convenience of use and the extent of their validation in empirical studies[17] makes members of our Cobb-Douglas family of special interest, and it is for these reasons that we have here focused on these types of functions. It should be noted, moreover, that our formulations extend to multiple outputs and to efficiency frontiers which are piecewise Cobb-Douglas. Also, we did not restrict the choice of μ_{ri} to the case of constant returns to scale, and, indeed, we may have increasing, decreasing and constant returns to scale occurring simultaneously on the same frontier.

In the cases of increasing, decreasing, constant or mixed returns to scale, we should perhaps specifically note that the efficiency evaluation is effected by reference to frontier segments. These and the inefficient ones they determine should be constrained to have the same properties. This could be done by introducing inequalities on the exponents or on the functions without altering the above models, and we emphasize again that these multiple output/multiple input formulations permit these features to be used in *some* of the output/input relations without requiring them for all of the others at the same time.[18] In any case, a gain in efficiency is always possible, at least in

[17] See Walters [17].

[18] Cf., e.g., (14) and the preceding remarks on the z_{r0}^* and z_{rj}^* which appear in it.

principle, for any DMU that is not operating on the frontiers for all of its inputs and outputs. This opens possibilities which may be used to evaluate "programs" as well as DMU's and their managers in ways that have been examined elsewhere. See [3], [4], and [9].[19]

[19]This is a revised version of an ealier paper entitled "A Bi-Extremal Principle for Estimating Efficiency Frontier Parameter Values". See [11]. Acknowledgment is due to Arie Lewin and to the referees for suggestions which guided parts of these revisions.

This research was partly supported by NSF Grant No. SOC76-15876, "Collaborative Research on the Analytical Capabilities of a Goals Accounting System." It is also supported by Project NR047-021, ONR Contracts N00014-75-C-0569 and 0616 with the Center for Cybernetic Studies, The University of Texas and ONR Contract N00014-76-C-0932 at Carnegie-Mellon University School of Urban and Public Affairs. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

References

1. BANKER, R., *Studies in Cost Allocation and Efficiency Evaluation*, DBA Thesis. Harvard University, Graduate School of Business Administration, Cambridge, Mass., 1980.
2. ———, "A Game Theoretic Approach to Measuring Efficiency," *European J. Operational Res.*, Vol. 5, No. 4 (1980), pp. 262–266.
3. CHARNES, A. AND COOPER, W. W., "An Efficiency Opening for Managerial Accounting in Not-for-Profit Entities," *Management Accounting* 1980, *Proceedings of a Conference on Research in Managerial Accounting*, Edited by H. Peter Holzer, University of Illinois, Department of Accountancy, Urbana, Ill., 1980.
4. ——— AND ———, "Auditing and Accounting for Program Efficiency and Management Efficiency in Not-for-Profit Entities," *Accounting, Organizations and Society*, Vol. 5, No. 1 (1980), pp. 87–107.
5. ——— AND ———, *Management Models and Industrial Applications of Linear Programming*, Wiley, New York, 1961.
6. ——— AND ———, "Management Science Relations for Evaluation and Management Accountability," *J. Enterprise Management*, Vol. 2, No. 2 (1980), pp. 143–162.
7. ——— AND ———, "Programming with Linear Fractional Functionals," *Naval Res. Logist. Quart.* Vol. 9, Nos. 3 and 4 (1962).
8. ———, ———, LEWIN, A., MOREY, R., AND ROUSSEAU, J., "Efficiency Analysis with Non-Discretionary Resources," Research Report CCS 379, The University of Texas at Austin, Center for Cybernetic Studies, Austin, Texas, September 1980.
9. ———, ———, AND RHODES, E. "Data Envelopment Analysis as an Approach for Evaluating Program and Managerial Efficiency—with an Illustrative Application to the Program Follow Through Experiment in U.S. Public School Education," *Management Sci.*, Vol. 27, No. 6 (1981), pp. 668–697.
10. ———, ———, AND ———, "Measuring the Efficiency of Decision Making Units," *European J. Operational Res.* Vol. 2, No. 6 (1978). See also "Short Communication," *European J. Operational Res.*, Vol. 3 (1979), p. 339.
11. ———, ———, AND Schinnar, A., "A Bi-Extremal Principle for Estimating Efficiency Frontier Parameters Values," Research Report CCS 329: Center for Cybernetic Studies, The University of Texas, Austin, Texas, March 1979.
12. ———, ———, AND ———, "A Theorem on Homogeneous Functions and Extended Cobb-Douglas Forms," *Proc. Natl. Acad. Sci. USA*, Vol. 73, No. 10 (1976), pp. 3747–3748.
13. ———, ———, AND ———, "Cobb-Douglas Representations for Analytic Functions," Research Report, Fels Discussion Paper No. 138, University of Pennsylvania, January 1979.
14. ———, ———, AND ———, "Transforms and Approximations in Cost and Production Function Relations," Research ReportCCS 339: The University of Texas, Center for Cybernetic Studies, Austin, Texas, 1976, 1977 and 1979.
15. RHODES, E., "Data Envelopment Analysis and Related Approaches for Measuring the 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, Pittsburgh, Penn., 1978.
16. SIMON, H. A., "Technology and the Environment," *Management Sci.*, Vol. 19, No. 10 (1973), pp. 1110–1112.
17. WALTERS, A. A., *An Introduction to Econometrics*, Norton, New York, 1970.