MANAGEMENT SCIENCE
Vol. 30, No. 9, September 1984
Printed in U.S.A

SOME MODELS FOR ESTIMATING TECHNICAL AND SCALE INEFFICIENCIES IN DATA ENVELOPMENT ANALYSIS*

R. D. BANKER, A. CHARNES AND W. W. COOPER
School of Urban and Public Affairs, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213
Graduate School of Business, University of Texas, Austin, Texas 78712

In management contexts, mathematical programming is usually used to evaluate a collection of possible alternative courses of action *en route* to selecting one which is best. In this capacity, mathematical programming serves as a planning aid to management. Data Envelopment Analysis reverses this role and employs mathematical programming to obtain *ex post facto* evaluations of the relative efficiency of management accomplishments, however they may have been planned or executed. Mathematical programming is thereby extended for use as a tool for *control* and evaluation of past accomplishments as well as a tool to aid in *planning* future activities. The CCR ratio form introduced by Charnes, Cooper and Rhodes, as part of their Data Envelopment Analysis approach, comprehends both technical and scale inefficiencies via the optimal value of the ratio form, as obtained directly from the data without requiring *a priori* specification of weights and/or explicit delineation of assumed functional forms of relations between inputs and outputs. A separation into technical and scale efficiencies is accomplished by the methods developed in this paper without altering the latter conditions for use of DEA directly on observational data. Technical inefficiencies are identified with failures to achieve best possible output levels and/or usage of excessive amounts of inputs. Methods for identifying and correcting the magnitudes of these inefficiencies, as supplied in prior work, are illustrated. In the present paper, a new separate variable is introduced which makes it possible to determine whether operations were conducted in regions of increasing, constant or decreasing returns to scale (in multiple input and multiple output situations). The results are discussed and related not only to classical (single output) economics but also to more modern versions of economics which are identified with "contestable market theories."
(EFFICIENCY; TECHNICAL INEFFICIENCY; RETURNS TO SCALE; MATHEMATICAL PROGRAMMING; LINEAR PROGRAMMING)

1. Background

Charnes, Cooper and Rhodes (CCR) (1978a, 1979) introduced a ratio definition of efficiency, also called the CCR ratio definition, which generalizes the single-output to single-input classical engineering-science ratio definition to multiple outputs and inputs without requiring preassigned weights. This is done via the extremal principle incorporated in the following model:

$$\max h_0 = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}} \quad \text{subject to} \quad 1 \geq \frac{\sum_{r=1}^{S} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}}, \quad j = 1, \ldots, n, \quad \text{with} \tag{1}$$

$$u_r, v_i > 0,^{1} \qquad i = 1, \ldots m; \quad r = 1, \ldots s.$$

Here the $y_{rj}, x_{ij} > 0$ represent output and input data for decision making unit (DMU) j

*Accepted by Arie Y. Lewin; received December 11, 1981. This paper has been with the authors $4\frac{1}{2}$ months for 2 revisions.

[1] See Charnes, Cooper, Lewin, Morey, and Rousseau for an exact non-Archimedean expression of this "positivity" with necessary algebraic closure.

0025-1909/84/3009/1078$01.25

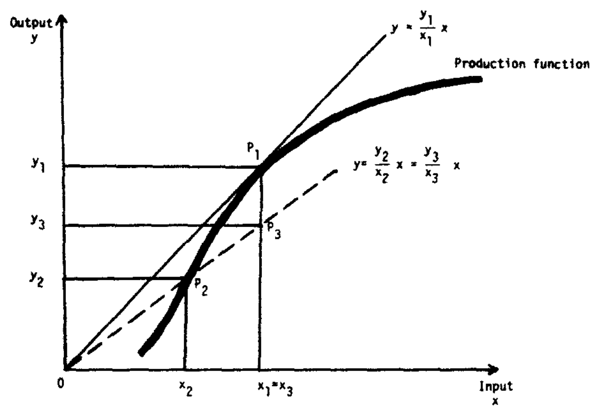


FIGURE 1

with the ranges for i, r and j indicated in (1). The data may be in the form of theoretically prescribed values or they may be in the form of observations. The unit to be rated is included in the functional with an index 0 as well as in the constraints, with the latter ensuring that an optimal $h_0^* = \max h_0$ will always satisfy $0 \leqslant h_0^* \leqslant 1$ with optimal solution values $u_r^*, v_i^* > 0$.

The main uses of these ideas have been in evaluations of "management" and "program" efficiencies[2] of decision making units (DMUs) of a not-for-profit variety such as schools,[3] hospitals,[4] etc. The ability to deal directly with multiple outputs and inputs forms one part of the appeal offered by these models and methods for uses such as these. Another part of its appeal comes from the development in Charnes, Cooper and Rhodes (1978a) which showed how the theory of fractional programming, as provided in Charnes and Cooper (1962), could be used to obtain access to a linear programming equivalent. This, in turn, yields an implementable form for securing solutions to (1) and it also yields a variety of duality relations for interpreting and utilizing the resulting $u_r^*, v_i^* > 0$.

Strong (and sharp) theoretical underpinnings as in physics and engineering are not available in applications such as we are considering. These must be replaced by weaker support—such as can be obtained from other disciplines like economics. It is, in fact, one purpose of the present paper to sharpen some of the latter contacts, but even after this has been accomplished, one must generally be satisfied with weaker results. For instance, one must be satisfied with a measure of only *relative* efficiency based on the available observations without recourse to what a stronger theory might provide.

We now try to clarify what has already been covered by reference to the illustration in Figure 1. Here we have portrayed the situation to be considered in terms of a single output, in amounts, y, and a single input, in amounts x. Three decision making units are to be rated for managerial efficiency. The production function represents the maximum output that can be produced for any specified input. The DMUs associated with P_2 and P_1 both achieve the maximum possible outputs for their input levels, while the DMU associated with P_3 falls short of the output level which is attainable from its x_3 input value.

To evaluate the efficiency of P_1, we utilize (1), which, for this one output-one input

[2] See Charnes, Cooper and Rhodes (1981), for further discussion of differences in "management" and "program" efficiency.

[3] See Bessent, Bessent, Kennington and Reagan (1982).

[4] See D. Sherman (1982) which also contains an interesting comparison that highlights deficiencies of statistical regressions (including translog and Cobb–Douglas regressions) and econometric estimation and similar approaches that have been addressed to these multiple output situations in the past.

case, becomes:

$$\max h_0 = \frac{uy_1}{vx_1} \qquad \text{subject to}$$
$$1 \geqslant uy_1/vx_1, \qquad 1 \geqslant uy_2/vx_2, \qquad 1 \geqslant uy_3/vx_3, \qquad u, v > 0, \tag{2}$$

where x_i, y_i represent the input and output coordinates of the DMU associated with P_i, $i = 1, 2, 3$. The ray from the origin tangential to the production function at P_1, lies above the ray through P_2 and P_3. This means that the DMU associated with P_1 is efficient and the other two are not. In fact, we have

$$u^* y_1 / v^* x_1 = 1 \qquad \text{with also} \tag{3.1}$$

$$u^* y_2 / v^* x_2 = u^* y_3 / v^* x_3 < 1. \tag{3.2}$$

This result is readily verified for this very simple case in which we evidently have

$$y_1/x_1 > y_2/x_2 = y_3/x_3. \tag{3.3}$$

See Figure 1. Thus for all u^*, $v^* > 0$ we also have the equality portrayed in (3.2) and the other conditions in (3.1) and (3.2) are necessary if u^*, v^* are optimizing solutions to (2).

This results in a characterization of the DMUs associated with P_2 and P_3 as being "equally inefficient" relative to the DMU associated with P_1. This characterization may be satisfactory in some cases. In other cases we may want to "fine tune" the developments in Charnes, Cooper and Rhodes (1978a) so that we can locate differences such as are portrayed in the P_2 and P_3 situations. Normally, of course, we will not have knowledge of the production functions but we can at least make a start toward this "fine tuning" to the extent that observational data may allow.[5] It is toward this end that we shall direct our proposed contacts with economics even while recognizing that the concepts and definitions of theoretical economics as formulated for applications to private sector market behavior may not always be best suited for management science (and related) applications in the not-for-profit sectors.

2. Production Technologies and Efficiency Envelopes

The economic theory of production forms a natural point of contact with economics and we elect to take this route via the concepts of R. W. Shephard. To be sure, Shephard's work (1953, 1970) is primarily directed toward developing formal relations between cost functions and a corresponding production technology on the assumption that a theoretically known efficiency has already been attained.[6] The production technology considered, however, encompasses the situation of multiple outputs in an unambiguous manner whereas other parts of (classical) production theory in economics are restricted to the single output situation—which is evidently of little or no interest in not-for-profit sector applications.[7] In addition, Shephard has introduced the concept of a "distance function"[8] which can be related to the important pioneering

[5] See Allen (1939) and Ferguson (1969) for treatments involving assumed knowledge of the production technologies. See also the survey by Kopp (1981).

[6] Note that economics concepts such as returns to scale, etc., have no unambiguous meaning until the efficiency frontier is attained. Thus, by virtue of this comment alone, most of the statistical-econometric studies on this topic are put in serious question. Other troubles may also be present as discussed in Charnes, Cooper and Schinnar (1982).

[7] Shortcomings of the classical single output theory even for purposes of economic theory are explained in Panzar and Willig (1977). See also Bailey and Friedlander (1982).

[8] Actually the distance function employed by Shephard is more properly regarded as a gauge function in the sense of Fenchel (1953).

work of M. J. Farrell[9] in measuring efficiency directly from observational data, at least in the single output case. In this way, as shown in Banker (1980b), Shephard's distance function[10] can be used to extend these ideas to more general situations, including those of a multiple output variety.

It should perhaps be explicitly observed that the CCR formulations in Charnes, Cooper and Rhodes (1981, 1978a) differ in important ways from the usual concepts of a production function at either the individual firm or aggregate level in both empirical and theoretical economics. In fact CCR refer to their function as an "envelope" developed relative to observational data from all of the $j = 1, \ldots, n$ DMUs, with the envelope forming an efficiency frontier relative to *each* firm (= DMU) that is to be evaluated. It is to be borne in mind, however, that it is not always appropriate to regard this envelope as a production function in the usual (classical) senses for some of the uses to which the CCR formulations may be put.

The developments we shall use here will be via the kinds of axiomatic formulations which have become common in this part of economics as a result of Shephard's work (1953, 1970). We shall try to do this, however, in a way that maintains contact with the kind of analytic formulations that are required for implementing these ideas in efficiency evaluations to be obtained directly from observational data.

3. Axioms

Our approach will be via optimizations conducted with respect to already generated observations.[11] For this situation we shall construct a simple model with input-output configurations observed for each of $j = 1, \ldots, n$ DMUs as (X_j, Y_j), where $X_j = (x_{1j}, \ldots, x_{ij}, \ldots, x_{mj})$ is a vector of observed inputs and $Y_j = (y_{1j}, \ldots, y_{rj}, \ldots, y_{sj})$ is a vector of observed outputs for DMU j. It is assumed that at least one output and at least one input are positive. Every DMU j used for efficiency comparisons is assumed to have used the same inputs and produced the same outputs, although, in general, in varying amounts. Our objective is to characterize a production possibility set and, in particular, to determine an "efficient" subset based on these observed data.

We shall represent the production possibility set as

$$T = \{(X, Y) \mid Y \geq 0 \text{ can be produced from } X \geq 0\}. \tag{4}$$

Following Shephard (1970, p. 179) we then define the input possibility set $L(Y)$, for each Y, as

$$L(Y) = \{X \mid (X, Y) \in T\} \tag{5}$$

and the output possibility set $P(X)$, for each X, as

$$P(X) = \{X \mid (X, Y) \in T\}. \tag{6}$$

We next postulate the following properties for the production possibility set, T:

Postulate 1. *Convexity*. If $(X_j, Y_j) \in T$, $j = 1, \ldots, n$, and $\lambda_j \geq 0$ are nonnegative scalars such that $\sum_{j=1}^{n} \lambda_j = 1$, then $(\sum_{j=1}^{n} \lambda_j X_j, \sum_{j=1}^{n} \lambda_j Y_j) \in T$.

Postulate 2. *Inefficiency Postulate*. (a) If $(X, Y) \in T$ and $\bar{X} \geq X$, then $(\bar{X}, Y) \in T$. (b) If $(X, Y) \in T$ and $\bar{Y} \leq Y$, then $(X, \bar{Y}) \in T$.

Postulate 3. *Ray Unboundedness*. If $(X, Y) \in T$ then $(kX, kY) \in T$ for any $k > 0$.

Postulate 4. *Minimum Extrapolation*. T is the intersection set of all $\hat{T}$ satisfying Postulates 1, 2 and 3, and subject to the condition that each of the observed vectors $(X_j, Y_j) \in \hat{T}$, $j = 1, \ldots n$.

[9]See Farrell (1957) and Farrell and Fieldhouse (1962).

[10]Shepard's distance function is also in need of fine tuning, as we shall see below, if it is to be used for efficiency characterizations. See also the discussion in Färe and Lovell (1978).

[11]In this respect, we follow Farrell (1957).

Thus, T is the "smallest" set consistent with the observed data and the postulated properties for the production possibility set. Because T is based on convexification and ray extension, it is a polyhedral set.

Next we seek to characterize any $(X, Y) \in T$. Postulates 1 and 3 imply that every (X, Y) of the form $(k\sum_{j=1}^{n}\lambda_j X_j, k\sum_{j=1}^{n}\lambda_j Y_j)$ with $k > 0$, $\lambda_j \geq 0$ and $\sum_{j=1}^{n}\lambda_j = 1$ is in T. Furthermore, employing Postulates 2 and 4, we can deduce that

$$\begin{aligned}&(X, Y) \in T \text{ if and only if } X \geq k\textstyle\sum_{j=1}^{n}\lambda_j X_j, \text{ and } Y \leq k\sum_{j=1}^{n}\lambda_j X_j,\\ &\text{for some } k > 0 \text{ and some } \lambda_j,\ j = 1, \ldots, n\\ &\text{satisfying the conditions } \lambda_j \geq 0 \text{ and } \textstyle\sum_{j=1}^{n}\lambda_j = 1. \end{aligned} \tag{7}$$

In the economics literature, Postulate 2 is sometimes referred to as "Free Disposability," a term which appears to have originated with Koopmans (1951). This may be a natural terminology for a discipline which has the analysis of market prices as a central concern. However, it risks a confounding of "market (allocative) efficiency" with the more basic condition of "technical efficiency" which is one of our central concerns.[12] Hence we prefer to use a term like "Efficiency Postulate" which we have converted to "Inefficiency Postulate" to indicate that inefficient production is always possible in the form of more inputs, smaller outputs, or both. In a similar way we prefer the more neutral term "ray unboundedness" to other alternatives such as "constant returns to scale." In this way we call attention to the fact that our usage of the latter is for individual characterizations and also to allow for other possibilities such as the presence of capacity limitations on inputs, especially when inefficiency possibilities are to be evaluated.

4. Shephard's Distance Function and Efficiency Measures

Having thus specified the production possibility set T, we next turn to the estimation of Shephard's distance function from these same observational data in order to relate it to the CCR efficiency measure. Shephard (1970, p. 206) defines a "distance function" $g(X, Y)$ of an input set $L(Y)$ as $g(X, Y) = 1/h(X, Y)$ where $h(X, Y) = \min\{h : hX \in L(Y), h \geq 0\}$.

Employing the characterization of any $(X, Y) \in T$ as in (4), we can express $h(X, Y)$ as:

$$\begin{aligned} h(X, Y) = \min h \quad &\text{subject to} \quad hX \geq k\sum_{j=1}^{n}\lambda_j X_j, \quad Y \leq k\sum_{j=1}^{n}\lambda_j Y_j, \\ &1 = \sum_{j=1}^{n}\lambda_j, \quad \lambda_j \geq 0 \quad \text{and} \quad k > 0. \end{aligned} \tag{8}$$

We next substitute $\mu_j = k\lambda_j$ in (8) above. Thus, we have

$$\begin{aligned} \min h \quad &\text{subject to} \quad hX - \sum_{j=1}^{n}\mu_j X_j \geq 0, \quad \sum_{j=1}^{n}\mu_j Y_j \geq Y, \\ &\mu_j \geq 0, \quad j = 1, 2, \ldots, n. \end{aligned} \tag{9}$$

This is a linear programming problem, for which a dual can be written as

$$\begin{aligned} \max U^T Y \quad &\text{subject to} \quad V^T X = 1, \quad U^T Y_j - V^T X_j \leq 0, \quad j = 1, \ldots, n, \\ &U \geq 0, \quad V \geq 0, \end{aligned} \tag{10}$$

where $U^T \equiv (u_1, \ldots u_r, \ldots, u_s)$ and $V^T \equiv (v_1, \ldots v_i, \ldots, v_m)$.

This is equivalent to the fractional programming problem with fractional con-

[12] The difference between these two types of efficiency are discussed in detail in Farrell (1957).

straints:

$$\max h = \frac{U^T Y}{V^T X} \qquad \text{subject to} \quad \frac{U^T Y_j}{V^T X_j} \leqslant 1, \qquad j = 1, \ldots n, \quad U, V \geqslant 0. \tag{11}$$

This could describe the CCR efficiency measure, except that each u_r and v_i in U and V is only required to be nonnegative rather than strictly positive. See (1).

The positivity requirement would be obtained, however, if we followed the non-Archimedean formulation and development in Charnes, Cooper and Rhodes (1979),[13] as we shall do below. Hence we shall say that we have established an equivalence between the CCR measure and the reciprocal of Shephard's distance function for input sets $L(Y)$, under the assumption that the production possibility set T satisfies the four postulates stated above.

Fine tuning is also required for Shephard's distance function, as already noted, if it is to be used for DMU efficiency measurement. Note, in particular, that the set $L_S(Y)$ defined by

$$L_S(Y) = \{X \mid h(X, Y) = 1\} \tag{12}$$

merely describes the boundary for inputs $X = (x_1, \ldots, x_m)$ that can be used to produce a given combination of outputs $Y = (y_1, \ldots, y_s)$. That is, following Shephard (1970, pp. 15–19) we allow for subsets of $L_S(Y)$ which we shall designate as

$$L_E(Y) = \left\{X \mid h(X, Y) = 1 \text{ with } \overline{X} \leqslant X \Rightarrow \overline{X} \notin L(Y) \text{ unless } \overline{X} = X\right\}. \tag{13}$$

This enables us to distinguish between points which are in the efficient subset $X_1 = (x_1, \ldots, x_i, \ldots, x_m) \in L_E(Y) \subseteq L_S(Y)$ and points which may be only on the boundary as in $X_2 = (x_1, \ldots, x_i + d, \ldots, x_m) \in L_s(Y)$ with $d > 0$.

The situation which enables us to distinguish between solutions like X_2 and X_1 in our linear programming formulation is the achievement of an optimum value of $h^* = 1$, but with positive slack in the ith input. In the situation for X_2, efficiency is not achieved despite an $h^* = 1$ since this inefficient consumption for input i may be reduced to the level X_i without affecting any other input or output.

We may clarify what is involved by introducing ideas from Charnes, Cooper and Rhodes (1978a). In particular, we replace (10) with the following problem:

$$\begin{gathered} \max z_0 = \sum_{r=1}^{s} u_r y_{r0} \qquad \text{subject to} \quad \sum_{i=1}^{m} v_i x_{i0} = 1, \\ \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leqslant 0, \qquad j = 1, \ldots, n, \quad \text{and} \quad u_r \geqslant \epsilon, \quad v_i \geqslant \epsilon \quad \forall r, i, \end{gathered} \tag{14}$$

where $\epsilon > 0$ is a small "non-Archimedean" quantity.

The dual to this problem is

$$\begin{gathered} \min w_0 - \epsilon\left[\sum_{i=1}^{m} s_i + \sum_{r=1}^{s} s_r'\right] \qquad \text{subject to} \\ 0 = w_0 x_{i0} - \sum_{j=1}^{n} x_{ij}\lambda_j - s_i, \qquad i = 1, \ldots, m, \\ y_{r0} = \sum_{j=1}^{n} y_{rj}\lambda_j - s_r', \qquad r = 1, \ldots, s, \qquad \lambda_j, s_i, s_r' \geqslant 0, \qquad \forall i, j, r. \end{gathered} \tag{15}$$

Evidently we can then have $z_0^* = w_0^* = 1$ if and only if the slack variables s_i and s_r' are all zero at an optimum.

REMARK. Färe and Lovell (1978) discuss the relationship between Farrell's efficiency measure and Shephard's distance measure. The results are ambiguous, however, because they begin by defining Shep-

[13] See also Charnes, Cooper, Lewin, Morey and Rousseau for a further refinement which relates the slacks of the ratio problem to the slacks of the linear programming problem.

hard's distance measure to be the Farrell efficiency measure in a way that attributes properties of the Shephard distance measure to the Farrell measure as well, and an error in their "proof" (p. 154) leads them to conclude that the reciprocal of Shephard's output distance function (rather than the distance function itself) corresponds to Farrell's efficiency measure. They also fail to credit Farrell with the use of points at infinity, which avoids the problems they address. The above, following Charnes, Cooper and Rhodes (1978a), replaces this with a simple formulation in which the non-Archimedean elements provide what is required in the constraints of one problem and in the functional of the corresponding dual.

To complete the cycle of connections between CCR efficiency and Shephard's distance function, with the related non-Archimedean extensions, we next turn to the output sets $P(X)$ as defined in (6). For this set, Shephard (1970, p. 207) defines a distance function

$$h'(X, Y) = 1/g'(X, Y) \qquad \text{where} \quad g'(X, Y) = \max\{ g' \mid g'Y \in P(X),\, g' \geqslant 0\}. \quad (16)$$

By following an analysis similar to the one we have used for the distance function for input sets, it is easy to establish a similar relation between Shephard's distance function and the CCR measure (before adjustment for slacks) when T satisfies Postulates 1 to 4.

5. Reduction of Postulates

In order to extend what we have already achieved to the task of tracing the efficient production surface, we now delete Postulate 3 from our requirements. This "Ray Unboundedness" postulate enabled us to extrapolate the performance of the most efficient DMUs with efficient scale sizes (for their given input and output mixes) and identify any scale inefficiencies that may be reflected in the level of operations of other DMUs. By deleting this postulate we now restrict our attention strictly to production inefficiencies at the given level of operations for each DMU, and thus develop an efficiency measurement procedure that assigns an efficiency rating of one to a DMU if and only if the DMU lies on the efficient production surface, even when it may not be operating at the most efficient scale size. This identification of the efficient production surface will also allow us to determine whether increasing, constant or decreasing returns to scale[14] prevail in different segments of the production surface. To reduce possible confusion with the concept of economies of scale, we shall avoid any recourse to cost functions[15] and related price imputation requirements and deal directly with the inputs and outputs as they may be observed in vectors X_j, Y_j for the jth DMU.

Assuming now that the possibility set T satisfies only Postulates 1, 2, and 4, we proceed as before to characterize T as the "smallest" set satisfying the "convexity" and the "inefficiency" postulates, subject to the condition that each of the observed vectors $(X_j, Y_j) \in T$. Therefore, proceeding as in the previous section, we can deduce that a vector (X, Y) is in the set T if and only if

$$X \geqslant \sum_{j=1}^{n} \lambda_j X_j, \qquad Y \leqslant \sum_{j=1}^{n} \lambda_j Y_j, \qquad (17)$$

for some $\lambda_j \geqslant 0$, $j = 1, \ldots, n$, satisfying the condition that $\sum_{j=1}^{n} \lambda_j = 1$.

We next determine Shephard's distance function for input sets $L(Y)$ when the production possibility set T is specified as above. Thus we have

$$g(X, Y) = 1/h(X, Y), \qquad \text{where} \quad h(X, Y) = \min\{h \mid hX \in L(Y),\, h \geqslant 0\} \quad (18)$$

which we translate as

$$\begin{gathered} \min h = h(X_0, Y_0) \qquad \text{subject to} \\ hX_0 - \sum_{j=1}^{n} \lambda_j X_j \geqslant 0, \quad \sum_{j=1}^{n} \lambda_j Y_j \geqslant Y_0, \quad \sum_{j=1}^{n} \lambda_j = 1, \quad \lambda_j \geqslant 0, \quad j = 1, \ldots, n, \end{gathered} \qquad (19)$$

[14] We are here referring to returns to scale in the restricted sense of theoretical economics. See §6 below.

[15] See, e.g., Farrell and Fieldhouse (1962) for further discussion.

relying on the fact that $h > 0$ will be satisfied when the components of every X_j and Y_j are all nonnegative—as is the case for the observational data we are considering.

This is a linear programming problem, the dual of which can be written as:

$$\max \sum_{r=1}^{s} u_r y_{r0} - u_0 \quad \text{subject to} \quad \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - u_0 \leq 0, \quad j = 1, \ldots, n,$$
$$\sum_{i=1}^{m} v_i x_{i0} = 1, \quad u_r, v_i \geq 0, \tag{20}$$

and u_0 is unconstrained in sign.

This linear programming problem is equivalent to a fractional programming problem,[16] which we express as follows

$$\max \frac{\sum_{r=1}^{s} u_r y_{r0} - u_0}{\sum_{i=1}^{m} v_i x_{i0}} \quad \text{subject to} \quad \frac{\sum_{r=1}^{s} u_r y_{rj} - u_0}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, \quad \forall j, \quad u_r, v_i \geq 0, \tag{21}$$

with, again, u_0 unconstrained in sign.

Similarly, Shephard's distance function for output sets $P(X)$ can be expressed as:

$$\max h'(X,Y) = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0} + v_0} \quad \text{subject to}$$
$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij} + v_0} \leq 1, \quad j = 1, \ldots, n, \quad u_r, v_i \geq 0, \tag{22}$$

and v_0 is unconstrained in sign.

Of course, in general, these two distance functions will not be identical. Furthermore, we observe (as in the preceding section) that the set $L_S(Y) = \{X \mid h(X,Y) = 1\}$, where $h(X,Y)$ is the reciprocal of Shephard's input distance function, describes only a (boundary) isoquant. This isoquant may not coincide with the efficient subset $L_E(Y) = \{X \in L(Y) \mid \bar{X} \leq X, \bar{X} \in L(Y) \Rightarrow \bar{X} = X\}$,

Following the analysis in the previous section, we introduce the infinitesimal non-Archimedean quantity[17] $\epsilon > 0$ to replace (20) with

$$\max \sum_{r=1}^{s} u_r y_{r0} - u_o \quad \text{subject to} \quad \sum_{i=1}^{m} v_i x_{i0} = 1, \tag{20A}$$
$$-\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} u_r y_{rj} - u_0 \leq 0, \quad j = 1, \ldots, n, \quad u_r, v_i \geq \epsilon \quad \forall r, i,$$

and u_0 is unconstrained in sign.

In a similar vein, the dual problem in (19) may be rewritten as

$$\min h - \epsilon\left[\sum_{i=1}^{m} s_i^+ + \sum_{r=1}^{s} s_r^-\right] \quad \text{subject to}$$
$$hx_{i0} - \sum_{j=1}^{n} x_{ij}\lambda_j - s_i^+ = 0, \quad i = 1, \ldots m,$$
$$\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^- = y_{r0}, \quad r = 1, \ldots s, \tag{19A}$$
$$\sum_{j=1}^{n} \lambda_j = 1, \quad \lambda_j, s_i^+, s_r^- \geq 0.$$

[16] We omit the transformation of variables implicit in this, as given explicitly in Charnes, Cooper and Rhodes (1978a), and, by abuse of notation, employ the same letters for the variables in both problems.

[17] As introduced in Charnes, Cooper and Rhodes (1978a, 1979).

6. Returns to Scale Characterizations

We now continue to extend our contacts with theoretical economics by turning to the problem of returns to scale (increasing, decreasing, or constant) as found in that literature. Once again, however, we emphasize that we are proceeding directly from observational data in our fine tuning of the CCR developments found in Charnes, Cooper and Rhodes (1978a). Mix and scale variations are likely to occur together in such observational data and so safeguards are required, as in our minimum extrapolation postulate, insofar as we cannot effect a separation of scale variations as in the formal theory of economics.[18] We shall examine returns to scale locally at a point, say (X_E, Y_E), on the efficient production surface, and relate it to the sign of the intercept term u_0 in the fractional programming problem (21) for this purpose.

We begin by asserting that the hyperplane given by

$$\sum_{r=1}^{s} u_r^* y_r - \sum_{i=1}^{m} v_i^* x_i - u_0^* = 0, \tag{23}$$

where the y_r, x_i are now variables, is a supporting hyperplane for the production possibility set T. Here u_r^*, v_i^*, and u_0^* are values of u_r, v_i, and u_0 that maximize the objective function in the fractional programming problem in (21).

We can confirm the support property of this hyperplane as follows. From the constraints to the problem in (21) we obtain

$$\sum_{r=1}^{s} u_r^* y_{rj} - \sum_{i=1}^{m} v_i^* x_{ij} - u_0^* \leqslant 0 \qquad \text{for} \quad j = 1, \ldots n. \tag{24}$$

Therefore, for any $\lambda_j \geqslant 0, j = 1, \ldots n$, with $\sum_{j=1}^{n} \lambda_j = 1$, we have

$$\sum_{r=1}^{s} u_r^* \sum_{j=1}^{n} y_{rj}\lambda_j - \sum_{i=1}^{m} v_i^* \sum_{j=1}^{n} x_{ij}\lambda_j - u_0^* \leqslant 0. \tag{25}$$

Further, by virtue of the result in (17), we can express any point $(X, Y) \in T$ as

$$\left(\sum_{j=1}^{n} \lambda_j X_j, \sum_{j=1}^{n} \lambda_j Y_j \right), \qquad \text{where} \quad \sum_{j=1}^{n} \lambda_j = 1, \quad \lambda_j \geqslant 0, \quad \text{all } j.$$

Therefore, we have

$$(X, Y) \in T \Rightarrow \sum_{r=1}^{s} u_r^* y_r - \sum_{i=1}^{m} v_i^* x_i - u_0^* \leqslant 0. \tag{26}$$

Also, since (X_E, Y_E) is efficient we have

$$\frac{U^{*T} Y_E - u_0^*}{V^{*T} X_E} = 1 \quad \text{or} \quad U^{*T} Y_E - V^{*T} X_E - u_0^* = 0. \tag{27}$$

Thus (26) and (27) together imply that $U^{*T} Y - V^{*T} X - u_0^* = 0$ is a supporting hyperplane for T at the point (X_E, Y_E).

Evidently this hyperplane is unique if and only if the optimal solution U^*, V^*, u_0^* to the equivalent linear programming problem is unique and this is the case we shall examine first. By virtue of (26) we have $U^{*T} Y - V^{*T} X - u_0^* \leqslant 0$ for all $(X, Y) \in T$. Hence in the unique supporting hyperplane situation which we are examining, a point (X_D, Y_D) in the neighborhood of (X_E, Y_E) will lie in the production possibility set if and only if $U^{*T} Y_D - V^{*T} X_D - u_0^* \leqslant 0$.

[18] See the discussion in Gold (1981) concerning the impracticality of effecting such separations from observational data so that characteristically the problem is usually approached from the cost function side. This, however, introduces problems with prices and allocative efficiencies.

To ascertain whether increasing, constant or decreasing returns to scale are present at (X_E, Y_E) we let $Z_\delta \equiv [(1+\delta)X_E, (1+\delta)Y_E]$ be a point in the neighborhood of (X_E, Y_E) by choosing δ to be suitably small. Then we can say that

(28a) Increasing returns to scale are present if and only if there exists a $\delta^* > 0$ such that (1) $Z_\delta \in T$ for $\delta^* > \delta > 0$ and (2) $Z_\delta \notin T$ for $-\delta^* < \delta < 0$.

(28b) Constant returns to scale are present if and only if there exists $\delta^* > 0$ such that (1) $Z_\delta \in T$ for all δ such that $|\delta| < \delta^*$, or (2) $Z_\delta \notin T$ for all δ such that $0 < |\delta| < \delta^*$.

(28c) Decreasing returns to scale are present if and only if there exists $\delta^* > 0$ such that (1) $Z_\delta \notin T$ for $\delta^* > \delta > 0$ and (2) $Z_\delta \in T$ for $-\delta^* < \delta \leqslant 0$.

Now,

$$U^{*T}(1+\delta)Y_E - V^{*T}(1+\delta)X_E - (1+\delta)u_0$$
$$= (1+\delta)\left(U^{*T}Y_E - V^{*T}X_E - u_0^*\right) + \delta u_0^* = \delta u_0$$

since we have $U^{*T}Y_E - V^{*T}X_E - u_0^* = 0$. Therefore, $Z_\delta \equiv ((1+\delta)X_E, (1+\delta)Y_E) \in T$ if and only if $\delta u_0 \leqslant 0$ since we are restricting our attention to the case of unique supporting hyperplanes at E for the polyhedral set T. Employing (28a), (28b) and (28c), it follows immediately that in the case when a unique supporting hyperplane passes through an efficient point (X_E, Y_E), we have

(29a) Increasing returns to scale $\Leftrightarrow u_0^* < 0$,

(29b) Constant returns to scale $\Leftrightarrow u_0^* = 0$,

(29c) Decreasing returns to scale $\Leftrightarrow u_0^* > 0$.

In other words, whether increasing, constant or decreasing returns to scale are present at (X_E, Y_E) depends on whether u_0^* is less than, equal to, or greater than zero in (20) and (21).

Figure 2 provides a generic picture of the situation in which scale changes are to be explored in the neighborhood of E. In the situation portrayed in Figure 2 we have $u_0^* > 0$ for the intercept value associated with the tangent line at E. The same situation would obtain for the succeeding piecewise segment on the efficient production possibility frontier. For the preceding segment where A appears, however, we would have $u_0^* < 0$ so that increasing returns would then be present. And, of course, for the

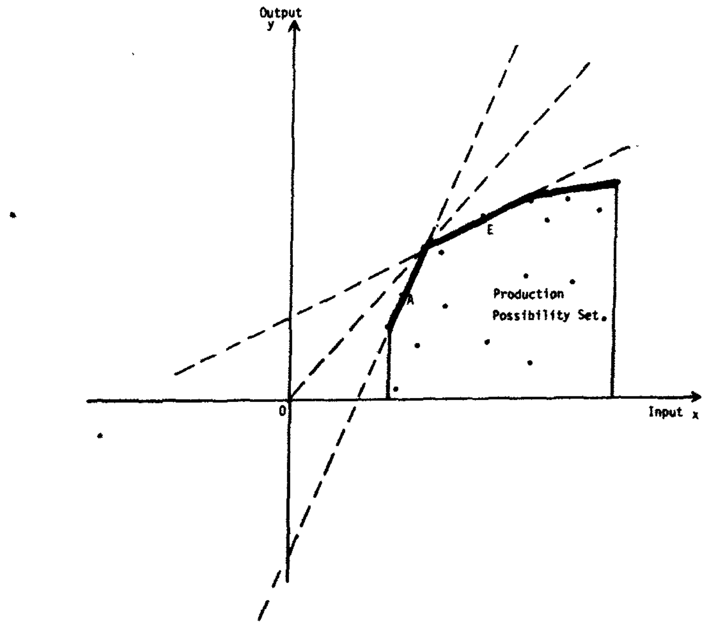


FIGURE 2

situation with $u_0^* = 0$ we would have constant returns if any such segment were present.

Up to this point we have been considering situations such as the one depicted generically at A and E in Figure 2 where the coefficients of the supporting hyperplane are determined by the components of the outward pointing normals at these points. At points of intersection such as shown for the two broken lines, these expressions are no longer unique. The components of the two normals at A and E provide alternate expressions, as do their convex combinations, from which additional supports may be designated that correspond to positions of rotation for these supporting hyperplanes at this point.

The above analysis considered efficiency measures based on input possibility sets as in (21). If we consider the alternative efficiency measures based on the output possibility sets, as in (22), then by pursuing a similar line of analysis we can relate the returns to scale at any point on the efficient production surface to the sign of the intercept term v_0^* for the supporting hyperplanes. In either case, in addition to providing a measure for the efficiency of individual observations and a means for tracing the efficient production surface, our mathematical programming formulation therefore also enables us to ascertain whether increasing, constant or decreasing returns to scale are present at specific points on the efficient production surface.

7. Concluding Interpretations

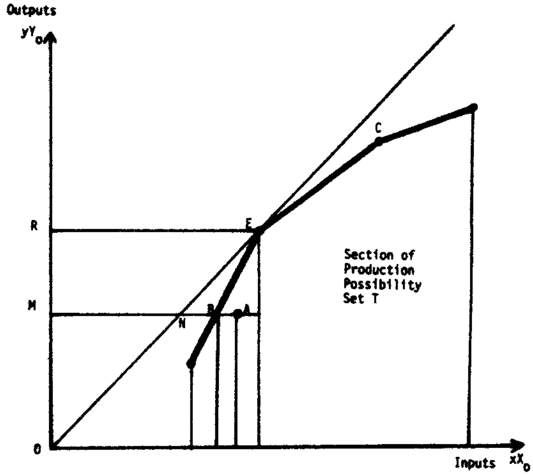
In this paper we have provided models for estimating technical and scale efficiencies of decision making units with reference to the efficient production frontier. The linear programming problems in (14) and (15) are employed to estimate the overall technical and scale efficiencies of a DMU. The linear programming formulations in (19A) and (20A) take into account the possibility that the average productivity at the most productive scale size may not be attainable for other scale sizes at which a particular DMU may be operating. These formulations estimate the pure technical efficiency of a DMU at the given scale of operation. The estimation of most productive scale size in DEA is discussed in Banker (1984).

Figure 3 illustrates these concepts of technical and scale efficiencies. The point A represents the DMU being evaluated. Its overall technical and scale efficiency is measured by the ratio MN/MA, by comparing the point A to the point N which reflects the average productivity attainable at the most productive scale size represented by the point E. The pure (input) technical efficiency of A is measured by the ratio MB/MA by comparing it with the point B on the efficient production frontier with the same scale size as A. Finally, the (input) scale efficiency of A is measured by the ratio MN/MB, so that the overall technical and scale efficiency MN/MA is equal to the product of the technical efficiency MB/MA and the scale efficiency MN/MB.

It is apparent from Figure 3 that the aggregate technical and scale efficiency measure MN/MA is less than the pure (input) technical efficiency measure MB/MA. This relationship between the two efficiency measures holds also for the general case of multiple inputs and outputs.

PROPOSITION. *The aggregate technical and scale efficiency measure as in* (15) *is less than or equal to the pure* (*input*) *technical efficiency measure as in* (19A), *with equality holding if and only if there exists an optimal solution to* (15) *such that the sum of the optimal values of the weights* $\lambda_j^*, j = 1, \ldots n$, *adds up to one. The proof follows immediately by comparing the constraint sets for the two formulations in* (15) *and* (19A).

These developments suggest that a scale efficiency measure can be defined as the ratio of the aggregate efficiency measure from (15) to the technical efficiency measure



Two-Dimensional Section of the Production Possibility Set T by the Plane Determined by $X = xX_0$ and $Y = yY_0$.

$A = (x_A X_0, y_A Y_0)$ represents the DMU being evaluated.

$B = (x_B X_0, y_B Y_0)$ represents a technically efficient referent point with the same (output) scale size.

$E = (x_E X_0, y_E Y_0)$ represents a technically and scale efficient referent point at the most productive scale size.

$$\text{(Input) Technical Efficiency} = \frac{MB}{MA} = \frac{y_A}{x_A} \Big/ \frac{y_B}{x_B} = \frac{x_B}{x_A}$$

$$\text{(Input) Scale Efficiency} = \frac{MN}{MB} = \frac{y_B}{x_B} \Big/ \frac{y_N}{x_N} = \frac{x_E}{x_B} \cdot \frac{y_B}{y_E}$$

$$\text{Technical and Scale Efficiency} = \frac{MN}{MA} = \frac{y_A}{x_A} \Big/ \frac{y_N}{x_N} = \frac{x_E}{x_A} \cdot \frac{y_A}{y_E}$$

FIGURE 3. Technical and Scale Efficiencies.

from (19A). Further the value of $\sum_{j=1}^{n} \lambda_j^*$ from (15) may be employed to measure the divergence from the most productive scale size. These possibilities are explored in greater detail in Banker (1984), and employed for a study of scale efficiencies and most productive scale sizes for hospitals in Banker et al. (submitted).

We shall not pursue these ideas any further. Instead we close by relating these ideas to recent developments in economics. This will help to highlight some of the shortcomings of the classical, single output, theory of production[19] and its public-policy consequences even in its classic (industrial) context while also highlighting some of the additional possibilities offered by DEA.

As a case in point we turn to J. C. Panzar and R. D. Willig (1977) who are able to call into question important results which have been standard for classical economics in areas such as the relations of economies to scale to conditions for monopoly growth and the use of marginal cost pricing without the need for subsidies and/or subventions to compensate for the losses that are supposed to accompany this approach to monopoly regulation under such conditions.

The developments in Panzar and Willig (1977) turn on the following formulation for determining whether returns to scale are present,

$$S = -\sum_{i=1}^{m} x_i \frac{\partial \phi}{\partial x_i} \Big/ \sum_{r=1} y_r \frac{\partial \phi}{\partial y_r}, \tag{30}$$

[19] See, e.g., Ferguson (1969).

with $S > 1$ if and only if increasing returns to scale[20] are present at (x, y). Here $\phi(X, Y) > 0$ is a known transformation function which is assumed to be differentiable at every point (X, Y) satisfying $\phi(X, Y) = 0$ where X and Y represent input and output vectors, respectively. Since attention is restricted to one firm at a time in Panzar and Willig (1977), no additional identifying subscript is needed. Finally, it is assumed that each such firm conducts all of its operations on the efficiency frontier.[21] How this is all to be implemented, tested or validated is not discussed in Panzar and Willig.

In our (DEA) approach we do not require knowledge of the transformation function and we also do not assume that each firm will attain the efficiency frontier. Instead, we employ the adjustment procedures described in Charnes, Cooper and Rhodes (1978a) in order to obtain new values $\hat{x}_{ij}$, $\hat{y}_{rj}$ which are all on the (relative) efficiency frontier for each of the $j = 1, \ldots, n$ "firms" (= DMUs) represented in the observations. With these thus adjusted input and output values we will then have

$$1 = \frac{\sum_{r=1}^{s} u_r^* \hat{y}_{rj} - u_{0j}^*}{\sum_{i=1}^{m} v_i^* \hat{x}_{ij}} \tag{31}$$

where we have indexed u_0^*, as from, e.g., repeated applications of (21), to provide the relevant identification for each of these $j = 1, \ldots, n$ firms.

This can be transformed straightforwardly into

$$\frac{\sum_{i=1}^{m} v_i^* \hat{x}_{ij}}{\sum_{r=1}^{s} u_r^* \hat{y}_{rj}} = 1 - \frac{u_{0j}^*}{\sum_{r=1}^{s} u_r^* \hat{y}_{rj}} = \hat{S}, \tag{32}$$

with $\hat{S} > 1$ if and only if $u_{0j}^* < 0$, i.e., if and only if increasing returns to scale are present, and $\hat{S} \leqslant 1$ otherwise. This then is our analogue for the result obtained by Panzar and Willig.

Note that we have not required their assumptions of differentiability and knowledge of the transformation function. Of course, if such knowledge is available then the measure of efficiency and adjustment to the frontier defined by that function can be effected in a straightforward manner.[22] In other cases, however, the DEA approach permits both efficiency measurement and adjustment to the relative frontier in a readily implemented manner.[23]

As Panzar and Willig note, their results apply only locally, and the same is true for $\hat{S}$. Here again, however, an extension has been effected and given operational form. Thus whereas the results in Panzar and Willig (1977) apply only at the point (X, Y), our results apply over the entire facet formed from the set of basis vectors used to evaluate DMU_j.[24]

Reference to (32) also provides insight into paths of possible further extension in these DEA approaches. Comparing the terms on the far left and far right, it becomes clear that our separation into scale and technical efficiency is obtained by relaxing the "unity-or-less" condition on the original CCR ratio formulation as given in (1) on p. 1078, above.

[20] Another approach to returns to scale which is suited to DEA and does not rely on the Panzar and Willig assumptions may be found in Banker (1984).

[21] It is assumed that the transformation function is also optimal since otherwise this firm may have its markets contested by new entrants. This all forms part of what is called "contestable market theory." As summarized in Baumol (1982), this theory retains the behavioral (optimizing) assumptions of the classical theory of competitive markets but (a) omits the requirement of having a large number of small firms and (b) switches the focus from single output to multiple output production.

[22] See the discussion in Charnes and Cooper (1980).

[23] A code for doing this and other parts of DEA which has been developed by I. Ali and J. Stutz is available from the Center for Cybernetic Studies, BEB200E, The University of Texas at Austin, Texas 78712.

[24] The computer code described in the preceding footnote identifies the facet and the basis from which it is formed.

Instead of pursuing the latter course,[25] we now close by simplifying the above analysis in a way that makes contact with classical, single output, production theory in economics. This can be done by first replacing the above transformation function with a production function $f(x)$ which is applicable in this single output case.

Removing the identifying subscript j, we write $y = f(x)$ for the output which is (uniquely) obtained from input x. Then we adjust the denominator in (32) for this single (efficient) output situation and obtain

$$\hat{S} = 1 - \frac{u_0^*}{y} = \frac{y - u_0^*}{y}. \tag{33}$$

If we further assume that x represents a single input then we can also write

$$y = x\frac{\partial f}{\partial x} + u_0^* \tag{34}$$

where $\partial f/\partial x$ and u_0^* represent the slope and intercept values associated with, e.g., the pertinent linear segment in Figure 3.

Substituting from (34) into the numerator on the right-hand side of (33) gives

$$\hat{S} = x\frac{\partial f}{\partial x} \Big/ y = \frac{\partial f}{\partial x} \Big/ \frac{y}{x}. \tag{35}$$

Thus we have increasing, constant or decreasing returns to scale according to whether

$$\frac{\partial f}{\partial x} \Big/ \frac{y}{x} \gtreqless 1 \quad \text{or equivalently,} \tag{36}$$

$$\partial f/\partial x \gtreqless y/x. \tag{37}$$

This is the well-known result in classical economics[26] that returns to scale are increasing, constant or decreasing at any pair of coordinate values (x, y) according to whether marginal product is greater than, equal to, or less than average product. In other words, all aspects of that theory are preserved when it is applicable. In other cases we may use DEA to augment or replace it without breaking contact with the underlying concepts.[27]

[25] One such possibility involves the potentially important "Congestion of Production Factors" introduced in Färe and Svensson (1980) as a possible source of inefficiency.

[26] See Allen (1939).

[27] The situation is not unlike the actual use of dual variables in linear programming in place of the *ceteris paribus* assumptions of classical economics which mainly serve as only guides to the concept of marginal opportunity cost because these conditions can only rarely be realized.

Parts of this paper rest on developments in R. D. Banker's doctoral thesis, "Studies in Cost Allocation and Efficiency Evaluation," at the Graduate School of Business of Harvard University, Boston, Massachusetts 02163, 1980. This research was supported in part by Project NR 1947-021, ONR Contract N00014-81-C-0236 with the Center for Cybernetic Studies, The University of Texas and ONR Contract N00014-81-C-0410 at Carnegie-Mellon University's School of Urban and Public Affairs. Reproduction in whole or in part is permitted for any purpose of the United States Government.

References

ALLEN, R. G. D., *Mathematical Analysis for Economists*, The MacMillan Co., New York, 1939.

BAILEY, E. E. AND A. F. FRIEDLANDER, "Market Structure and Multiproduct Industries," *J. Econom. Lit.*, 20 (September 1982), 1024–1048.

BANKER, R. D., "A Game Theoretic Approach to Measuring Efficiency," *European J. Oper. Res.*, (1980a), 262–266.

———, "Studies in Cost Allocation and Efficiency Evaluation," unpublished doctoral thesis, Harvard University, Graduate School of Business Administration, 1980. Also available from University Microfilms International, Ann Arbor, Mich., 1981.

———, "Estimating Most Productive Scale Size Using Data Envelopment Analysis," *European J. Oper. Res.*, (1984).

———, A. CHARNES, W. W. COOPER AND A. SCHINNAR, "A Bi-Extremal Principle for Frontier Estimation and Efficiency Evaluation," *Management Sci.*, 27 (December 1981).

———, R. CONRAD AND R. P. STRAUSS, "An Application of DEA to the Empirical Investigation of a Hospital Production Function," *Management Sci.* (submitted).

BAUMOL, W., "Contestable Markets: An Uprising in the Theory of Industry," *Amer. Econom. Rev.*, 72, 1 (March 1982), 1–15.

BESSENT, A., W. BESSENT, J. KENNINGTON AND B. REAGAN, "An Application of Mathematical Programming to Assess Productivity in the Houston Independent School District," *Management Sci.* 28, 12 (December 1982).

CHARNES, A. AND W. W. COOPER, "An Efficiency Opening for Managerial Accounting in Not-For-Profit Entities," in P. Holzer (Ed.), *Proceedings of a Conference on Managerial Accounting*, University of Illinois Department of Accounting, Urbana, 1980.

——— AND ———, "Managerial Economics: Past, Present and Future," *J. Enterprise Management*, 1 (1978).

——— AND ———, "Programming with Linear Fractional Functionals," *Naval Res. Logist. Quart.*, 9, 3, 4 (1962).

———, ———, A. LEWIN, R. MOREY AND J. ROUSSEAU, "An Approach to Positivity and Stability Analysis in DEA," *Ann. Oper. Res.* (submitted).

———, ———, AND E. RHODES, "Data Envelopment Analysis: Approach for Evaluating Program and Managerial Efficiency—with an Application to the Program Follow Through Experiment in U.S. Public School Education," *Management Sci.*, 27, 6 (June 1981), 668–697.

———, ———, AND ———, "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, (1978a), 429–444.

———, ———, AND ———, "Short Communication: Measuring the Efficiency of Decision Making Units," *European J. Oper. Res.*, (1979), 339.

———, ———, AND A. SCHINNAR, "Transforms and Approximations in Cost and Production Function Relations," *Omega*, 10, 2 (1982), 207–211. See also accompanying "Prologue."

———, ———, AND L. SEIFORD, "An Empirical DEA Production Function," 1981. Reissued as CCS Report 459 "Pareto-Optimality, Efficiency Anaysis and Empirical Production Functions," by A. Charnes, W. W. Cooper, B. Golany, L. Seiford and J. Stutz, University of Texas at Austin, 1983.

FÄRE, R. AND C. A. K. LOVELL, "Measuring the Technical Efficiency of Production," *J. Econom. Theory*, (1978), 150–162.

——— AND L. SVENSSON, "Congestion of Production Factors," *Econometrica*, (November 1980), 1745–1753.

FARRELL, M. J., "The Measurement of Productive Efficiency," *J. Roy. Statist. Soc.* Ser. A, 120 (1957), 253–290.

——— AND M. FIELDHOUSE, "Estimating Efficient Production Frontiers under Increasing Returns to Scale," *J. Roy. Statist. Soc.* Ser. A, (1962), 252–267.

FENCHEL, W., *Convex Sets, Cones and Functions*, Princeton University, Department of Mathematics, Princeton, N.J., 1953.

FERGUSON, C. E., *The Neoclassical Theory of Production and Distribution*, Cambridge University Press, Cambridge, 1969.

FØRSUND, F. R., C. A. K. LOVELL AND P. SCHMIDT, "A Survey of Frontier Production Functions and of their Relationship to Efficiency Measurement," *J. Econometrics*, (1980).

GOLD, B., "Changing Perspectives on Size, Scale and Returns: An Interpretive Survey," *J. Econom. Lit.*, (March 1981).

KOOPMANS, T., *Activity Analysis of Production and Allocation*, John Wiley & Sons, Inc., New York, 1951.

KOPP, R. J., "The Measurement of Productive Efficiency: A Reconsideration," *Quart. J. Econom.*, (1981), 477–503.

PANZAR, J. C. AND R. D. WILLIG, "Economies of Scale in Multi-Output Production," *Quart. J. Econom.*, (August 1977), 481–493.

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. Public School Education," Ph.D. Thesis, Carnegie-Mellon University, School of Urban and Public Affairs, Pittsburgh, 1978. Also available from University Microfilms International, Ann Arbor, Mich., 1979.

SHEPHARD, R. W., *Cost and Production Functions*, Princeton University Press, Princeton, N.J., 1953.

———, *The Theory of Cost and Production Functions*, Princeton University Press, Princeton, N.J., 1970.

SHERMAN, D., "Data Envelopment Analysis and Other Approaches for Measuring the Efficiency of Hospital Operations," DBA Thesis, Harvard University Graduate School of Business, 1981. Also available from University Microfilms International, Ann Arbor, Mich., 1982.