This article was downloaded by: [130.39.62.90] On: 20 August 2014, At: 12:11
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through

A. Charnes, W. W. Cooper, E. Rhodes,

To cite this article:
A. Charnes, W. W. Cooper, E. Rhodes, (1981) Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through. Management Science 27(6):668-697. http://dx.doi.org/10.1287/mnsc.27.6.668

Full terms and conditions of use: http://pubsonline.informs.org/page/terms-and-conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

© 1981 INFORMS

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

RIGHTSLINK

MANAGEMENT SCIENCE
Vol. 27, No. 6, June 1981
Printed in U.S.A

EVALUATING PROGRAM AND MANAGERIAL EFFICIENCY: AN APPLICATION OF DATA ENVELOPMENT ANALYSIS TO PROGRAM FOLLOW THROUGH*

A. CHARNES,† W. W. COOPER† AND E. RHODES‡

A model for measuring the efficiency of Decision Making Units (= DMU's) is presented, along with related methods of implementation and interpretation. The term DMU is intended to emphasize an orientation toward managed entities in the public and/or not-for-profit sectors. The proposed approach is applicable to the multiple outputs and designated inputs which are common for such DMU's. *A priori* weights, or imputations of a market-price-value character are not required.

A mathematical programming model applied to observational data provides a new way of obtaining empirical estimates of extremal relations—such as the production functions and/or efficient production possibility surfaces that are a cornerstone of modern economics. The resulting extremal relations are used to envelop the observations in order to obtain the efficiency measures that form a focus of the present paper.

An illustrative application utilizes data from Program Follow Through (= PFT). A large scale social experiment in public school education, it was designed to test the advantages of PFT relative to designated NFT (= Non-Follow Through) counterparts in various parts of the U.S. It is possible that the resulting observations are contaminated with inefficiencies due to the way DMU's were managed en route to assessing whether PFT (as a program) is superior to its NFT alternative. A further mathematical programming development is therefore undertaken to distinguish between "management efficiency" and "program efficiency." This is done via procedures referred to as Data Envelopment Analysis (= DEA) in which one first obtains boundaries or envelopes from the data for PFT and NFT, respectively. These boundaries provide a basis for estimating the relative efficiency of the DMU's operating under these programs. These DMU's are then adjusted up to their program boundaries, after which a new inter-program envelope is obtained for evaluating the PFT and NFT programs with the estimated managerial inefficiencies eliminated.

The claimed superiority of PFT fails to be validated in this illustrative application. Our DEA approach, however, suggests the additional possibility of new approaches obtained from PFT-NFT combinations which may be superior to either of them alone. Validating such possibilities cannot be done only by statistical or other modelings. It requires recourse to field studies, including audits (e.g., of a U.S. General Accounting Office variety) and therefore ways in which the results of a DEA approach may be used to guide such further studies (or audits) are also indicated.

(PROGRAM EFFICIENCY; MANAGERIAL EFFICIENCY; EFFICIENCY FRONTIERS)

1. Introduction

This paper is concerned with evaluating the efficiency of public programs and how they are managed. It is directed to ways by which program efficiency might be disentangled from management efficiency by reference to empirical observations obtained from the organizations (e.g., schools) or other decision making

*Accepted by Arie Y. Lewin. Received on March 13, 1979, this paper was refereed as an unusual contribution under the guidelines that were applicable at that time. It has been with the authors 9 months for 2 revisions.

†University of Texas at Austin.
‡State University of New York at Buffalo.

0025-1909/81/2706/0668$01.25
Copyright © 1981, The Institute of Management Sciences

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

units which convert *designated* program inputs into *desired* program outputs. Our procedures for effecting such efficiency evaluations will be illustrated by reference to data selected from the results of a large-scale social experiment in public school education for disadvantaged children known as Program Follow Through.[1]

Some such disentanglement is needed if we are to avoid imputing the results of good management to bad programs and *vice versa*. For public sector evaluations, moreover, the likelihood of confounding is increased by the presence of multiple outputs and inputs. Handling such programs as single outputs via subjective weighting schemes introduces a degree of arbitrariness that is unacceptable for many applications. The same is true even when objective data are used to effect a reduction from multiple outputs to a single output via cost-price imputations effected by reference to markets that are often remote from the programs being considered.

We might, as an illustrative case in point, refer to "self-esteem" as one of 11 desired outputs, and to "parental attention to children" as one of 25 inputs designated as important for evaluating the results of the Program Follow Through Experiment. How or why these particular outputs and inputs were determined to be important will not be examined in this paper.[2] We shall instead assume that the desired outputs and the designated inputs, as well as the way they are to be measured, have already been determined. This will allow us to focus on the issue of efficiency in the conduct of such programs which we may characterize by reference to the following output and input orientations:

i. *Output Orientation*: A Decision Making Unit (= DMU) is not efficient if it is possible to augment any output without increasing any input and without decreasing any other output.

ii. *Input Orientation*: A DMU is not efficient if it is possible to decrease any input without augmenting any other input and without decreasing any output. (1)

A DMU will be characterized as efficient if, and only if, neither (i) nor (ii) obtains.

As we shall see, our measure of efficiency allows us to take account of both of the above orientations at the same time. For clarity of presentation, however, it seemed best to separate them for the purposes of the immediately following discussion. In any case, the above characterizations may be regarded as extending the usual concept of "Pareto efficiency"[3] by including inputs (as well as outputs) on the supposition that any released inputs have "some" value in other possible uses. As such, our measure is free of the arbitrariness of the weighting and price imputation approaches referred to above, at least as far as efficiency is concerned. We do not want to stop, however, with only an identification of efficient and inefficient DMU's. We want to go still further and provide estimates of the amount of efficiency as well as the dimensions (programs or managerial) in which they occur.

The term Decision Making Unit (= DMU) is introduced in the above definitions in order to avoid the strictures that long usage has accorded such terms as "plant" and/or "firm" as organization entities concerned with input and output decisions in the literature of economics. We shall find it helpful, however, to borrow from that literature in order to use the concept of a "production function" for clarifying aspects of what we are trying to accomplish with our efficiency evaluations. This is what we now proceed to do in the remainder of this introductory section. That is, we use the idea of a production function to provide previews of the concepts and methodologies, as well as the applications that will be effected later in this article.

[1] Also called The Follow Through Planned Variation Experiment, and sometimes referred to as Project Follow Through, or the Follow Through Project or Program. Exhibit A in the Appendix provides information on the scale of this experiment. Exhibits B and C provide information on the portion used for the study reported in this paper.

[2] These topics are discussed in [1] and [2]. See also [43] and [53]. As Eric Hanushek in [44] and [45] observes, such parental and family inputs have been generally regarded as important—perhaps of paramount importance—ever since the Coleman report.

[3] Also called Pareto-Koopmans efficiency. See [17, Chapter IX]. In an alternate approach, which he does not explicitly relate to P-K efficiency, Farrell [36] refers to what we are here concerned with as "technical efficiency" which he distinguishes from "price efficiency." We do not deal with such price efficiency considerations in this paper. The means for doing so are available, however, via the values of the marginal rates of substitution which, as we shall see, are available as byproducts for use along the efficiency frontiers via the estimating procedures we will employ. See [28] and [50] for further discussion. The extensions of transform theory described in [30] are available if the related (efficiency) cost functions are also wanted.

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

To commence with these characterizations, we write

$$y = f(x_1, \ldots, x_m), \tag{2}$$

wherein y represents the level of a single output obtained from any specified vector of input values via $x = (x_1, \ldots, x_m)$. For (2) to represent a "production function," the value of y must be maximal for each such collection of input values. In other words, f is an "extremal relation." It is defined by reference to a given technology (i.e., state of managerial and engineering knowledge) which includes the way decision makers within the DMU are (or should be) organized.[4]

Because (2) is applicable to every one of the $j = 1, \ldots, n$ "firms" within an "industry," we may use it to evaluate the efficiency of any one along the following line. Suppose we are given the input values $x_{1j}, \ldots, x_{mj}$ for the jth such firm. Substitution in (2) would yield

$$\hat{y}_j = f(x_{1j}, \ldots, x_{mj}) \tag{3}$$

as the maximal output obtainable from these input values. If the observed output value is y_j, we would then have

$$0 < y_j/\hat{y}_j \leq 1 \tag{4}$$

to provide a measure of the efficiency achieved by this firm. This measure is more than just an index of efficiency. It is wholly operational in the sense that its value represents the proportion of the possible output that was actually obtained—or, *per contra*, it represents the loss of output from anything less than an efficient utilization of these same input values.[5]

Under the circumstances indicated above, such efficiency discrepancies can be attributed to management.[6] Of course, the situation may be more complex in that some managers may be required to operate with production functions that differ from the others. We might then designate $\hat{y}_j^\alpha$ as the maximal output which is possible for each of $\alpha = 1, 2, \ldots, k$ different production functions and similarly index the input vectors $x_j^\alpha = (x_{1j}^\alpha, x_{2j}^\alpha, \ldots, x_{mj}^\alpha)$ in order to evaluate the efficiency of management under each of these k different sets of production possibilities.

The introduction of such different sets of production possibilities raises efficiency questions in its own right. Even assuming that all managers operate at 100% efficiency, one might want to know which of these functions is most efficient, and how this kind of efficiency might be identified and evaluated.

This type of "functional efficiency,"[7] which involves a comparison between functions, brings us to the distinction between "program efficiency" and "managerial efficiency" that is examined later in the paper. The former is measured by reference to managerial behavior under its appropriate function or program. The latter involves an across-program comparison between different functions. As such, it involves additional complications even when it is supposed that differences arising from managerial efficiency have all been identified and allowed for. Notice, for instance, that one production function may admit greater output values than another over some ranges of input values but the reverse may be true over other input ranges. The situation is, of course, even more complicated when more than two functions are involved. Hence, although our treatment will allow for extensions to this class of cases, we will not treat them in detail in the present paper.

Of course, the fact that one function is more efficient than others only over some ranges of input values can be important in its own right. This would be the case, for example, when institutional arrangements for the resource allocation decisions make it possible to consider their assignment to different firms or programs —as when, for instance, plant locations in different regions, or legal restrictions associated with different programs, are pertinent.

Our analysis allows for such identifications when pertinent. In some cases of public policy evaluation, however, we may want to consider how best to choose one from among the various possible functions. Such might be the case when a decision must be made to continue only one of several federal government programs, e.g., as in the case of Program Follow Through, even though one program is more efficient under some conditions while the situation is reversed under other conditions.

[4] See Carlson [15] and/or Allen [5] for a discussion of the way organization considerations—i.e., the way decisions are organized and related to each other—enter into the definition of a production function.

[5] The development here follows the one used in [27]. See also [38].

[6] We leave aside the possibility of observational and other such errors. See Aigner *et al.* [3]. The kinds of statistical developments needed to deal with the kinds of distributions associated with our proposed efficiency measures are discussed in [29] and will not be treated here.

[7] This usage of the term "functional efficiency" differs from the one on pp. 321ff. in [17].

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

To guarantee that the efficiency ratio which will be used for such an evaluation cannot exceed unity, we erect a new function which is always at least as efficient as any function in the set. This function is referred to as the "inter-envelope" in the text that follows since it envelops the other functions in the set. The latter may also be regarded as envelopes, in the form of "efficiency frontiers" which provide bounding conditions for the DMU's operating under them. Hence, the name Data Envelopment Analysis (= DEA) which arises from the procedures (and concepts) applied to observational data which are used to establish the efficiency frontiers via these envelopment procedures.

Supposing that the managers of these DMU's could previously move only to the boundaries defined by their respective production functions—or program boundaries—we now suppose that they can move their DMU's to the boundary defined by this new inter-envelope. On this supposition, the resulting measure of efficiency then has properties which extend the previous definition of efficiency. Our approach takes account, in particular, of the distribution of DMU's under each program, and the resulting augmentations and/or resource conservation possibilities that can be achieved by movement from the old to the new inter-envelope boundary for each pertinent program.

The following sections of this paper examine possible ways for accomplishing these purposes. This is done by turning the machinery of mathematical programming from its usual planning (only) uses in management decision making in order to apply these "optimizing" tools and concepts to an evaluation of already effected decisions.[8] This results in new ways to estimate relations—e.g., extremal relations—from empirical observations. Comparisons with other approaches such as fitting statistical regressions to empirical observations are introduced, not only as a way to facilitate interpretation of our suggested new approaches, but also as a way to indicate when this approach—or the other more customary ones in statistics—may best be used. Combinations of the two approaches are also possible, of course, and will be indicated. The paper's emphasis is on efficiency measurement, however, and not on estimation of these extremal relations.[9]

2. Decision Making Efficiency

If we were concerned with only one program, we might import the following formulation from [28] and [50] as a measure of managerial (= decision making) efficiency:

$$\max h_0 = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}}$$

$$\text{subject to } 1 \geq \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}}; \qquad j = 1, \ldots, n; \tag{5}$$

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

Here, as before, the x_{ij} represents input values for the jth DMU, but now we index the outputs so that y_{rj} represents the observed amount of each of $r = 1, \ldots, s$ outputs obtained from these inputs. All of the outputs and inputs are assumed to be observed as positive values. Each of these $j = 1, \ldots, n$ DMU's utilizes the same inputs and produces the same outputs in (generally speaking) different amounts. We may, in fact, say that this suffices to characterize them as belonging to the same "industry."

The n constraints in (5) evidently ensure that no DMU can achieve an efficiency rating which will exceed unity. For each phase of an efficiency evaluation, one member of this set is singled out and represented in the functional, as well as in the constraints. The ensuing optimization then yields a *positive* set[10] of u_r^*, v_i^*

[8] This differs from uses in which results from programming (planning) models are compared with past decisions and, on the basis of past data, the results are compared with those of the mathematical programming models to see whether significant improvements might have been attained. See, e.g., [25] and [46]. See also Theil [55].

[9] See [28] and [50] for further discussion.

[10] The reason for restricting these to positive values is set forth in the Corrections to [28]. See also the Non-Archimedean Model Efficiency Theorem in [24] which relates the presence of positive slack in the linear programming problem to the presence of Non-Archimedean elements in the solution of the ratio problem.

RIGHTSLINK

Efficiency Points and Isoquant.

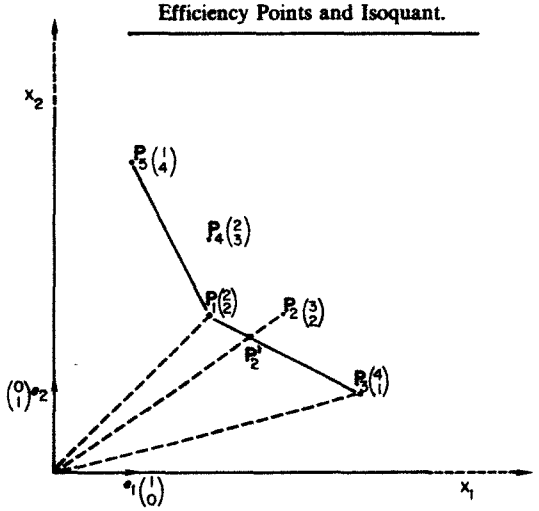


FIGURE 1.

which generate an optimal $0 < h_0^* < 1$ with $h_0^* = 1$ if, and only if, the thus distinguished DMU is efficient in the sense of (1).

We now proceed to interpret these h_0^* and u_r^* and v_i^* values by relating them (in a linear programming "duality" sense) to concepts in production economics as follows. Suppose, for the moment, that we are again in the single output (ordinary production function) situation. Observations on two inputs, in the amounts x_1 and x_2, have been secured for each of 5 DMU's. These input values are then normed on their respective outputs and exhibited as the points $P_1, \ldots, P_5$ relative to the coordinate axes indicated in Figure 1 so that the coordinates in Figure 1 represent the different inputs as rates per unit output. In the case of P_2, for example, the input rates per unit output are $x_{12} = 3$ and $x_{22} = 2$, respectively, where the index j (here $j = 2$) indicates which of the 5 DMU's is being referenced for each of the $i = 1, 2$ inputs.

In the interest of simplicity, we shall refer to these x_{ij} values in terms of "amounts" rather than rates. Then we shall refer to the solid line in Figure 1 as the "unit isoquant" which we will use for efficiency evaluations in the following manner.

Consider the point P_2 with the indicated coordinates. These are the observed inputs required by this DMU to produce one unit of output. On the other hand, the point P_2' on the ray from the origin to P_2 is on the unit isoquant. But this means that P_2' also produces one unit of this same output with less of each of the same inputs. Hence, on the assumption that it is meaningful to do so, we may use the ratio

$$0 < d(OP_2')/d(OP_2) < 1 \tag{6}$$

as a measure of efficiency for this DMU, where $d(OP_2')$ represents the length of the ray from the origin to P_2' and $d(OP_2)$ represents the length of this same ray from the origin to P_2.

Evidently the resulting value would be less than unity for P_2 and the same would be true in a similar measurement applied to the data associated with P_4. This same approach would, however, produce a value equal to unity for each of the other 3 DMU's portrayed in Figure 1. The latter would thus all be characterized as being efficient, by virtue of being on the unit isoquant, while this would not be true for P_2 and P_4. We have thus obtained on the dual programming side a characterization of efficiency, which, by duality, is equivalent to our ratio formulation. See [24] and [28].

The details for doing this have been explained in [28][11] and we will not repeat them here. Instead, we simply record the optimal values obtained for the evaluation of the DMU associated with P_2 as

$$u^* = 1, v_1^* = \tfrac{1}{6}, v_2^* = \tfrac{1}{3} \tag{7}$$

which gives an optimal value for the functional

$$h_0^* = \frac{1u^*}{3v_1^* + 2v_2^*} = \frac{6}{7} \tag{8.1}$$

[11] See also Rhodes [50].

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

and satisfies all the pertinent constraints, *viz.*,

$$\begin{aligned}
\frac{1u^*}{2v_1^* + 2v_2^*} &= 1, \\
\frac{1u^*}{3v_1^* + 2v_2^*} &= \frac{6}{7}, \\
\frac{1u^*}{4v_1^* + 1v_2^*} &= 1, \\
\frac{1u^*}{2v_1^* + 3v_2^*} &= \frac{3}{4}, \\
\frac{1u^*}{1v_1^* + 4v_2^*} &= \frac{2}{3},
\end{aligned} \tag{8.2}$$

as required for (5) in this simple unit amount of a single output case.

Evidently P_2 is not efficient and, in fact, the $h_0^* = \frac{6}{7}$ means that it should have been able to achieve its one unit of output while using $\frac{6}{7}$ of the amount of each of the observed inputs. This is established by reference to the data for P_1 and P_3 which are designated as efficient via this same calculation, as can be seen from values in the first and third constraints in (8.2).

What this example shows is that although we have made *no* prior assumptions about the production structures of the DMU's, linear programming duality associates with each ratio formulation a "tangent" polyhedral or constant-returns-to-scale production frontier for each DMU which is a *local* approximant to its production function.

Although our measure of efficiency appears to be relative only to subsets of the DMU's being evaluated, it must be borne in mind that all the others have participated in characterizing their efficiency as unity. Thus the $h_0^* = \frac{6}{7}$ efficiency value for P_2 was obtained by the ray process relative to P_1 and P_3 where, on the basis of the data in Figure 1, the latter were characterized as efficient for purposes of this evaluation.

We can further clarify what we are saying by turning to the evaluation of P_4 which, via the route we have just indicated,[12] gives

$$u^* = 1, v_1^* = \tfrac{1}{3}, v_2^* = \tfrac{1}{6} \tag{9}$$

for which we obtain

$$h_0^* = \frac{1u^*}{2v_1^* + 3v_2^*} = \frac{6}{7} \tag{10.1}$$

This is the same efficiency value obtained for DMU_2. This ray process efficiency evaluation for DMU_4 is by reference to P_5 and P_1, however, as may be verified by substitution in the pertinent constraints from (5) to obtain

$$\begin{aligned}
\frac{1u^*}{2v_1^* + 2v_2^*} &= 1, \\
\frac{1u^*}{3v_1^* + 2v_2^*} &= \frac{3}{4}, \\
\frac{1u^*}{4v_1^* + 1v_2^*} &= \frac{2}{3}, \\
\frac{1u^*}{2v_1^* + 1v_2^*} &= \frac{6}{7}, \\
\frac{1u^*}{1v_1^* + 4v_2^*} &= 1.
\end{aligned} \tag{10.2}$$

This differs from the situation for the $h_0^* = \frac{6}{7}$ obtained for DMU_2 where the ray efficient referent set was provided by P_1 and P_3. In other words, the ray efficient reference sets for P_2 and P_4 differ even though the application of the resulting efficiency value may be given the same resource shrinkage interpretation in both cases.

[12] I.e., by proceeding via a linear programming equivalent to obtain the optimal values for use in (5).

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

When approaching the issue of efficiency from an isoquant standpoint, it is natural to speak in terms of input reductions. We can also turn the matter around to a consideration of output augmentation and, indeed, for the kinds of models we are using, our measures may be applied to outputs and inputs simultaneously, if desired.[13]

A consideration of how this may be done can be obtained by reference to the solution values in (7) and (9) in a way that can supply additional illumination. The v_1^* and v_2^* values there, as obtained from the linear programming solutions, represent the normals to the corresponding isoquant segments in Figure 1. This segment stretching from P_1 and P_3 represents the set defined by

$$\{(x_1, x_2) : \tfrac{1}{6}x_1 + \tfrac{1}{3}x_2 = 1; 2 \leq x_1 \leq 4; 1 \leq x_2 \leq 2\}, \tag{11}$$

where the coefficient values are obtained from (7) via $v_1^* = \frac{1}{6}$ and $v_2^* = \frac{1}{3}$. These are evidently the marginal productivities assigned to these input variables. They are not, however, the marginal productivities that might be estimated from data on the activities of P_2 only. They are, instead, the *efficient* marginal productivities applicable for the evaluation of any DMU which can be referred to this isoquant segment. Hence, no j subscript appears as a DMU indicator for either of the variables x_i, $i = 1, 2$ in (11).

The range in which these productivities apply (also available from the tableaus) are shown in the right-hand portion of (11). Within this range, the indicated relation is linear so that marginal and average productivities are equal there. Hence, applying these v_i^* values to the observed values for P_2 we have

$$v_1^* x_{12} + v_2^* x_{22} = 3v_1^* + 2v_2^* = \tfrac{7}{6}. \tag{12}$$

We thereby arrive at an output value of $\frac{7}{6}$ units (the reciprocal of the efficiency ratio) instead of the one unit of output that P_2 secured.

The corresponding productivities for the isoquant segment between P_1 and P_5 are secured from (9) so that the resulting set is represented by

$$\{(x_1, x_2) : \tfrac{1}{3}x_1 + \tfrac{1}{6}x_2 = 1; 1 \leq x_1 \leq 2; 2 \leq x_2 \leq 4\}. \tag{13}$$

This differs from (11), of course, although a similar interpretation by reference to the data again produces

$$\tfrac{1}{3}x_{14} + \tfrac{1}{6}x_{24} = \tfrac{7}{6}. \tag{14}$$

Although the output augmentation is the same in both cases, it is important to bear in mind that these values are obtained from different efficient reference sets. Thus, it would be erroneous to apply the productivities in (13) to P_2 and it would also be erroneous to apply the productivities in (11) to estimate the output augmentation that is possible from efficient utilization of the inputs utilized by P_4.

To underscore what is involved, we should emphasize that our production surface is *not* linear. It is only *piecewise* linear, with the result that the productivities are also only piecewise constant and applicable only within specified ranges before *different* efficient productivities become applicable.

It is legitimate, of course, to inquire whether a still more general representation than the ones provided by this piecewise linear approach might be utilized. Reference [7], for instance, shows how to extend these ideas to functions which are piecewise Cobb-Douglas, also with multiple outputs, as well as even more general classes of functions. Recourse to this approach would admit analyses of situations such as those involving increasing and decreasing (as well as constant) returns to scale in different outputs at the same time.

For the application we shall shortly be considering, however, it seems precipitous to extend the boundaries of the possible analyses in this manner. The scoring of outputs like a child's self-esteem, or inputs like parental attention is not at a stage where it would be easy to distinguish an actual appearance of such phenomena as increasing returns to scale from deficiencies of the scoring mechanisms. Under such circumstances, it seems best to proceed as we have always done[14] by starting from simpler models such as the piecewise linear one we are using in the present article while allowing for further step-by-step progress—as the discovery of new knowledge may suggest in the course of such applications.

There is a further advantage in following such a source in that computer codes have been devised for

[13] See [35] for further discussion of the conditions under which this can be done.

[14] E.g., in the first actual industrial use of linear programming where we proceeded via the linearization allowed by "performance numbers" used to rate TEL mixtures for gasoline blending purposes instead of proceeding immediately to the nonlinearities associated with direct uses of octane number ratings. See [25]. Also see the exchange between A. Charnes and A. Manne in this same volume of *Econometrica*. Subsequent developments, of course, have since accommodated these and other nonlinear relations in the mathematical programming models that are now almost universally used in gasoline and other blending operations.

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

conducting the desired efficiency calculations and these codes also provide the productivities and other coefficients estimates, etc., as desired, for further use in applications and extensions.[15] These codes comprehend the case of multiple outputs and inputs included in (5), naturally, and so it seems desirable to conclude this section of the paper by extending our previous interpretations to this class of situations as well.

An extension beyond the single output case requires a similar extension of the production function concept to the more general case of "production possibility surfaces" and their related efficiency properties.[16] Concepts like marginal productivity, and related isoquant facets, etc., become ambiguous in such multiple output situations, and *a fortiori*, this is true when these outputs are simultaneously determined from whatever input combination might be used. This, in fact, is the class of cases with which we are concerned, and so we proceed to an interpretation of the following kind.

Borrowing from natural science terminology in physics and engineering,[17] we shall refer to the numerator values for the functional in (5) as providing a "virtual input" and the denominator as providing a "virtual output." The u_r and v_i will be regarded as "virtual rates of transformation" from the observed input and output values into these virtual inputs and outputs. The result will be a dimensionless ratio which transforms into the actual (efficiency measure) ratio when these u_i^* and v_i^* are optimal. The v_i^* are interpretable as marginal productivities relative to this virtual input as a measure of "output potential." Similarly, for any such output potential as may be specified,[18] the u_i^* may be interpreted as efficient marginal rates of transformation from the observed outputs to the virtual output with a resulting output-to-input ratio that will not exceed unity.

In conclusion, we might observe that our measure of efficiency is scale independent in each of its inputs and outputs. Thus, if we have an h_0^* obtained from one set of x_{ij} values, then replacing any subset of these x_{ij} with new values $\rho_i x_{ij}$, $\rho_i > 0$, in both the functional and the constraints will not alter the original h_0^*. The same property holds for the outputs y_{rj}, or for any combinations of inputs and outputs that might be of interest provided these changes of scale apply to all of the DMU's being considered.[19]

3. Managerial Efficiency

We now want to extend the above ideas to enable us to distinguish program from managerial efficiency in the different reference sets of DMU's we shall be studying. We therefore introduce the following extension of (5):

$$\max h_0^\alpha = \frac{\sum_{r=1}^{s_\alpha} u_r^\alpha y_{r0}^\alpha}{\sum_{r=1}^{m_\alpha} v_i^\alpha x_{i0}^\alpha}$$

$$\text{subject to } 1 \geq \frac{\sum_{r=1}^{s_\alpha} u_r^\alpha y_{rj}^\alpha}{\sum_{i=1}^{m_\alpha} v_i^\alpha x_{ij}^\alpha}; \qquad j = 1, \ldots, n_\alpha \tag{15}$$

$$u_r^\alpha, v_i^\alpha > 0; \qquad r = 1, \ldots, s_\alpha; \quad i = 1, \ldots, m_\alpha,$$

where $\alpha = 1, 2, \ldots, k$, respectively, indexes the different sets which are of interest.

Within each set we will, of course, have the same efficiency measurement situation as before—*viz.*, $0 < h_0^{*\alpha} \leq 1$ with $h_0^{*\alpha} = 1$ if and only if the DMU being evaluated relative to the αth set of DMU's is efficient. Now, however, we want to extend these ideas so that we can apply them across the sets $\alpha = 1, 2, \ldots, k$ in order to examine the relative efficiency of the associated programs.

For this comparison, we shall require common outputs and inputs for the reference sets. Then, provisionally, we may think of this as a comparison between each of $\alpha = 1, 2, \ldots, k$ "technologies" in order

[15] See [47].

[16] See [28].

[17] See the discussion on pp. 645ff. in [17] on the programming uses of such concepts as "virtual displacement", "virtual work," etc.

[18] See R. G. D. Allen [5] for a discussion of the economic significance of the usual marginal rates of transformation from one output to another.

[19] See [18] for a proof. Other, more general, invariance properties are examined in [50].

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

to determine their varying degrees of efficiency for converting common inputs into common outputs. Each such technology provides a "boundary" to the set of production possibilities under the usual assumptions of economic theory. We shall be dealing with inferences from empirical data, however, in which we will *not* be able to assume that all DMU's attain these boundaries. Generally having no knowledge of these boundaries from some *a priori* source, we shall only be able to establish them by an envelopment procedure, as explained in Figure 1, by reference to the most efficient members of the respective sets of DMU's. These efficient subsets of DMU's will then be used to establish the *relative* efficiency boundaries which we shall refer to as "envelopes" in order to emphasize these (and other) departures from the usual assumptions of economic theory and methods of empirical inquiry.

In any case, we shall use $\alpha = 1, 2, \ldots, k$ such envelopes for measuring the efficiency of the DMU's operating under each such set. Then we shall be concerned with adjusting each such set of DMU's onto its pertinent envelope prior to making across-envelope efficiency comparisons in the manner indicated in the opening section of this paper. To effect these across-envelope comparisons, we shall bring each DMU onto the envelope for its reference set in the manner set forth in [28]. We shall mainly be concerned with behavior, such as the behavior of educational institutions, in the public sector—where perfectly competitive market forces are not ordinarily given free play. Nevertheless, in a rough sort of analogy, we may think of these intra-envelope adjustments as corresponding to that part of competitive theory in which each DMU is forced to become as efficient as the most efficient of its competitors as a condition for survival.[20] Note, however, that this is not an assumption in our case since we actually adjust the observations in this manner. Thus we are able to effect these across-envelope comparisons on the basis of data which are adjusted so that all DMU's are as efficient as the most efficient among them. Comparisons across these envelopes will then be used to rate the respective efficiencies of these envelopes.

Statistical aspects of the testing and estimation procedures that might be employed have been treated elsewhere—see [18] and [50]—and will not be discussed in any detail in the present paper. We might, however, underscore points like the following. Unlike the situation in which one wants to test an underlying theory, we are here proceeding in a manner which uses that theory to bring the observations onto the envelope that serves as the efficiency frontier in each set. Only *after* this has been done are the tests of significance to be applied for the inter-envelope comparisons.

This approach considerably simplifies some of the statistical models and methods that may be employed, and it opens a variety of applications for policy evaluations and controls that are not available from more customary approaches.[21] In any event, we need to distinguish this approach which we shall refer to as Data Envelopment Analysis (DEA) and which we now try to motivate in the following way: Suppose we have two different programs that might be used in public education. Each program has the same (multiple) outputs and utilizes the same inputs as, for instance, in the experiment on Program Follow Through (PFT) that we shall shortly examine. In deciding whether PFT is better than its Non-Follow Through (NFT) alternative, we need to allow for a variety of possibilities in view of the fact that the observations for each of PFT and NFT contain deviations that can reflect decisions by management which fall short of what each *program* admits.

By distinguishing between program and managerial (= decision making) efficiency, our DEA approach is directed toward evaluating a variety of policy possibilities that need to be considered. As already noted, it is directed to distinctions like those between managerial and program efficiency so that, *inter alia*, we can determine whether program comparisons entail different degrees of managerial efficiency in the data sets, or whether allowances should be made for different degrees of DMU efficiency before effecting program evaluations. Furthermore, the DEA approach singles out the more efficient DMU's for possible study *en route* to setting standards and other types of controls within any such program. It also opens the possibility of synthesizing entirely new programs by identifying subsets of across-program DMU's as a possible source for forming new program combinations that are better than any of the originally identified programs. Of course, still other possibilities become available and, in any case, the DEA approach helps us to distinguish good programs which might be badly managed from worse programs that appear to be better because of management rather than program capability. It is this latter aspect of DEA which we shall emphasize in what follows, but we shall also at least indicate some of these other possibilities along the way.

[20] We shall also refer to our envelopes as "efficiency frontiers" even though we do not make the usual profit maximizing (incentive) assumption that the most efficient DMU's always effect the best choice that technology makes possible.

[21] For a discussion of the meager results obtained to date from the usual econometric approaches to education see Griliches [42].

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

4. Program Follow Through Background

We shall illustrate these DEA ideas by reference to a body of data[22] that has recently become available from a very important experiment in U.S. public school education known as Project Follow Through (PFT).[23] Before commencing with the specifics of our analysis, however, we briefly consider the history and development of the Project Follow Through experiment.

It was conceived in the late 1960's as a federally sponsored program charged with providing remedial assistance to educationally disadvantaged early primary school students. To a large extent PFT was developed in response to perceived needs for furthering the objectives and accomplishments of the well known Project Head Start.[24] In fact, a major justification for Follow Through was to supplement Head Start as a pre-school program, since the public schools did not articulate sufficiently with the "Head Start goals, curricula, and objectives in the early grades to enable these children to maintain or accelerate their pre-school achievement."[25] The Follow Through Program was envisioned as an answer to the latter problem. At the same time, it was also to be a "community action program," going well beyond the classroom in providing for community services such as nutrition programs, social, medical, and dental assistance, and even psychological counseling service.

The academic portion of Follow Through could be interpreted as a form of Head Start which moved the latter from pre-school into the elementary grades at the level of kindergarten through third grade. Initially conceived as a program involving some 200,000 children, Follow Through received enthusiastic endorsements from a variety of educational authorities. Unfortunately for those anticipating a large-scale primary school action program, a series of developments occurred between Follow Through inception and funding which resulted in a change of federal policy and a reduction in funding.[26] This caused a rethinking in which the proposed massive application was converted into an "experimental study" with the latter to be executed in an approach referred to as "planned variation."

The idea was to utilize an experimental design approach or at least as much of an approximation to these canons of classical statistics as one is likely to be able to secure in a field like educational policy.[27] Within the limits of the "planned variation model," the Follow Through Program was to be formulated around a collection of specifically identified approaches to treating the compensatory education problem of disadvantaged children. These program variations were each associated with "sponsors" (e.g., sponsors headquartered at local universities, or research institutes) who were to (1) provide the basic form and content of one particular "planned variation" and (2) work with designated local school districts in implementing the indicated variation. See Appendix, Exhibit B.

Conformance with the above conditions was to be a requirement for federal funding (and related resource advantages) and, further, this was extended to a directive that each school district supply a Non-Follow Through as well as a Follow Through candidate group. Allowance was made for periodic reports and analyses to facilitate study of these various programs, and competent statistical (and other) consultants were retained for effecting analyses of the resulting data. The results from these analyses were so mixed and

[22] Actually, our analysis is based on data made available to us through the courtesy of the U.S. Office of Education and Abt Associates, Inc., of Cambridge, Massachusetts—who made these data available to us even in advance of the analyses presented in [1] and [2]. This is an unusual courtesy, especially for data on public school education, which we herewith gratefully acknowledge. We are also grateful to Mary Kennedy, Program Follow Through Project Officer at the U.S. Office of Education, for sending us the U.S. Office of Education Reports that are also referenced in [2] and [43].

[23] The study itself is known as *Education as Experimentation: A Planned Variation Model*. See the discussions in the U.S. Office of Education reports listed in [2] and [43].

[24] Project Head Start was designed as an early childhood *pre-school* intervention program aimed at bringing about significant cognitive and non-cognitive gains among disadvantaged children. When subsequent studies indicated that Project Head Start effects were not sustained after its participants entered primary school, Project Follow Through was one suggested corrective measure—the idea being that special attention in the first few grades would lend reenforcement to what Project Head Start had previously initiated. See the discussions on pp. 158–159 in Vol. IIA of the U.S. Office of Education Reports in [2].

[25] From [53, pp. 2–3].

[26] See Appendix, Exhibit A, for indications of the scope of the Project Follow Through Experiment.

[27] This is, of course, only a recognition of the particular susceptibility of education, and especially education in the early grades, to emotions, pressures and other impediments to purely scientific studies. See Haney [43].

RIGHTSLINK

subject to dispute and challenge, however, that we confine ourselves to a PFT versus NFT comparison without reference to the variations in the assorted Follow Through approaches identified with these different sponsors.[28]

Another difficulty arises in that Follow Through provided a variety of social and medical, as well as educational services to the community. Being community based—e.g., because of its tie to the Community Action Program arm of the U.S. Office of Economic Opportunity—and not being attached to specific academic programs, it was not possible to determine the differential effects, if any, of these activities of Follow Through. Thus, like other analysts, we will simply ignore these parts of the Program in order to focus on only the academic portions of Follow Through.

Within the latter limits, certain attractive features emerge for our purposes. For one thing, the Follow Through study is almost unique among programs of its size in that all of the sites administered the same core battery of tests and measurements for the proposed national evaluation. This included the NFT as well as PFT segments. Moreover, the former, i.e., the NFT sites, were selected to obtain matched comparison sets of supposedly comparable students. The PFT results could thereby be matched to comparable control populations rather than being confined only to comparisons with some aggregate national norm. While this matching was not completely carried out in all detail, it at least provides a better basis than most of the other "quasi-experimental" designs[29] of "planned variation" genres in educational policy.

5. Selection of Variables

We have already indicated some of the properties of our proposed DEA approach to program and managerial efficiency measurement. Now we indicate others. Note, for instance, that the above matching presents certain difficulties that are not encountered in the classical (natural science) models of experimental design and which therefore require specific attention. As a case in point, we might consider the problem of managerial (= decision making) efficiency as it might be distributed between DMU's in PFT and NFT. Differences in decision making efficiency need to be allowed for since, evidently, a "good program" may be "badly managed," and *vice versa*, so that one needs some way of identifying this possible source of contamination in arriving at a "program" evaluation.

If these were profit making entities one might—at least in principle—use dollar scalarizations for both inputs and outputs in order to effect a matching for efficiency, possibly in the original experimental design. No such *a priori* basis is available here, however, and so our DEA procedures are applied "after the fact," so to speak, to eliminate such managerial inefficiencies en route to effecting the wanted PFT-NFT comparisons.

Since we want to focus on the concepts and adjustment methodologies associated with our DEA approach, it seems prudent to restrict ourselves to only a few of the variables for which data are available from the PFT experiment. This means that our application to Program Follow Through is only illustrative. On the other hand, the variables we shall study are important ones and so our adverse findings, even though used mainly to illustrate the uses of DEA, cannot be simply brushed aside. Moreover, omitted parts of the program (such as the community services components) should have biased the results in favor of PFT. In other words, even a favorable outcome for PFT would have fallen short of what is required is that further justification for these other expenditures and activities would be needed before a positive recommendation for PFT was warranted.[30] The fact that our study is not favorable to PFT compared to NFT means that strong effects in other dimensions are needed to compensate for this.[31]

[28] See [1] and [2] for a detailed treatment of the various Follow Through sponsor performances.

[29] In the sense of Campbell and Stanley [13].

[30] Note, however, that our proposed approach can accommodate the usual dollar scalarization of a cost-benefit analysis as only one (or a subset) of the many dimensions which should enter into an efficiency evaluation, instead of according these dollar scalarizations over-riding importance, as in the more customary cost-benefit approaches.

[31] Actually, the separation between PFT and NFT is not as complete as might be desired. For one thing, other Title I experiments might have been underway in some of the NFT components and possible contaminating effects could also emerge from even social interchanges between NFT and PFT participants. See pp. 13ff. in Vol. II-A of U.S. Office of Education [2]. There are other difficulties, too, of a technical nature. See, e.g., Ferber and Hirsch [37] who, p. 91, cite Alice Rivlin on the added difficulty of estimation problems in social experiments concerned with altering functional forms in order to increase the efficiency of, e.g., education processes. Note, however, that we are attempting to achieve the latter in a different manner by (a) fixing the functional forms, with reference to efficient DMU's, and then (b) bringing the other

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

We shall focus on only one of several cohorts from the subjects comprehended in the study.[32] In addition, we shall utilize only the terminal (grade 3) results for this cohort to avoid the additional complications needed to deal with dynamic or transient behavior en route to this terminus. From a set of 11 output measures, we select only the following 3 as sufficiently indicative for our purposes:

y_1: Total Reading Score as measured by the Metropolitan Achievement Test.[33]

y_2: Total Mathematics Score as measured by the Metropolitan Achievement Test.[33]

y_3: Coopersmith Self-Esteem Inventory, intended as a measure of self-esteem.[34]

This y_3 measure, we may note, is directed to affective behavior (or noncognitive growth) in a dimension that was deemed pertinent to the objectives of this program. Together with y_1 and y_2, this y_3 variable provides a good indication of what is involved in assessing such programs. Note, in particular, that no easily available scheme for weighting the relative importance of these outputs is at hand, and even the notion of a market for effecting imputations of the value of self-esteem to disadvantaged young children seems artificial or even bizarre. Nevertheless, some "overall" measure of program efficiency is wanted in order to enable us to evaluate PFT vs. its NFT alternative and this is to be achieved from data such as are exhibited in Tables 1 and 2 (for PFT) and Tables 3 and 4 (for NFT).

The latter table, i.e., Table 2, contains the input data for those same PFT sites.These input values were selected from among a set of 25 as follows:

x_1: Education level of mother as measured in terms of percentage of high school graduates among female parents.

x_2: Highest occupation of a family member according to a pre-arranged rating scale.

x_3: Parental visit index representing the number of visits to the school site.

x_4: Parent counseling index calculated from data on time spent with child on school-related topics such as reading together, etc.

x_5: Number of teachers at a given site.

All of the output data in Tables 1 and 3 are in units of 100 students and the same is true for the input data of Tables 2 and 4 with the exception of x_5.[35] Note that we again have a fair indication of what is likely to be encountered in such studies in that only x_5 offers any real prospect for the kinds of input variation decisions that are customary in the economic theory of production. To be sure resources may be spent to encourage parental visits or counseling, but such controls are loose at best, and even the number of teachers may admit of only limited variation.

As already noted, we can nevertheless proceed to evaluate "efficiency" by reference to whether the maximal outputs have been achieved from the inputs utilized—at least after we have decided upon the

DMU's up to the resulting surface. See [20] and [21] for yet another treatment of such experimental data as a decision problem rather than an estimation problem—in the context of marketing new products, including the choice of studies used to guide a product's design components and marketing strategies.

[32] As observed in note 3, Exhibit C of the Appendix.

[33] See [49].

[34] See [34].

[35] See [26] and [50] for further discussion and elaboration of the reasons for these choices. As already noted, the value of h_0^* as a measure of relative efficiency is not affected by such scale choices provided different scales are not used at different sites for any particular input or output. See [18] and [50].

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

TABLE 1

Unadjusted PFT *Output Observations*

Site #	Total Reading Scores Y_1	Total Math Scores Y_2	Total Coopersmith Scores Y_3
1	54.53	58.98	38.16
2	24.69	33.89	26.02
3	36.41	40.62	28.51
4	14.94	17.58	16.19
5	7.81	6.94	5.37
6	12.59	16.85	12.84
7	17.06	16.99	17.82
8	20.29	30.64	33.16
9	26.13	29.80	26.29
10	46.42	51.59	35.20
11	39.80	37.73	30.29
12	37.84	47.85	25.35
13	26.48	31.36	26.54
14	10.31	10.86	7.47
15	14.39	18.30	14.33
16	32.94	36.03	38.19
17	17.25	20.80	12.07
18	27.55	38.19	20.44
19	41.12	43.80	36.54
20	29.43	42.63	23.34
21	37.46	51.02	27.44
22	19.40	25.18	16.52
23	39.88	47.72	38.97
24	25.72	30.81	16.54
25	24.88	25.27	22.43
26	31.62	40.78	31.16
27	31.31	38.32	25.03
28	21.00	21.30	18.30
29	6.51	7.02	6.16
30	11.64	15.26	15.68
31	12.58	15.90	14.42
32	4.59	6.16	4.99
33	43.76	46.64	39.10
34	32.38	38.55	31.05
35	34.64	45.46	39.22
36	11.52	15.14	13.91
37	15.96	19.21	15.30
38	9.91	12.30	7.22
39	30.44	33.53	29.80
40	22.63	25.24	17.15
41	24.41	27.16	25.30
42	23.11	22.67	17.56
43	21.82	31.45	27.54
44	63.92	79.67	63.11
45	9.47	11.92	8.85
46	33.94	39.18	34.61
47	29.42	35.10	28.42
48	7.70	11.02	9.02
49	12.17	16.03	15.82

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

TABLE 2

Unadjusted PFT *Input Observations*

Site #	Education Level of Mother X_1	Occupation Index X_2	Parental Visit Index X_3	Counseling Index X_4	Number of Teachers X_5
1	86.13	16.24	48.21	49.69	9
2	29.26	10.24	41.96	40.65	5
3	43.12	11.31	38.19	35.03	9
4	24.96	6.14	24.81	25.15	7
5	11.62	2.21	6.85	6.37	4
6	11.88	4.97	18.73	18.04	4
7	32.64	6.88	28.10	25.45	7
8	20.79	12.97	54.85	52.07	8
9	34.40	11.04	38.16	42.40	8
10	61.74	14.50	49.09	42.92	9
11	52.92	11.67	39.48	39.64	5
12	36.00	10.15	37.80	39.52	5
13	39.20	10.80	41.04	41.12	7
14	14.60	2.88	9.64	11.14	3
15	4.29	5.42	21.45	17.27	5
16	27.25	14.17	56.46	55.26	9
17	22.63	4.43	15.40	15.00	2
18	28.00	7.61	28.73	27.04	9
19	53.56	13.70	53.04	49.85	7
20	25.42	9.05	29.69	31.74	4
21	31.57	10.08	39.34	40.57	6
22	16.34	5.84	20.89	22.10	4
23	44.28	14.14	56.70	52.27	11
24	19.74	6.43	24.20	25.66	3
25	24.40	8.05	33.42	31.29	7
26	41.40	11.70	44.01	46.35	7
27	27.20	9.38	37.80	31.55	4
28	23.92	7.12	25.58	29.01	3
29	10.62	2.55	10.10	9.09	4
30	12.48	6.14	23.13	22.46	6
31	19.32	5.89	24.01	24.74	6
32	6.30	1.93	7.11	7.68	4
33	46.62	14.65	65.71	57.49	10
34	38.95	12.82	47.02	48.92	9
35	61.60	15.56	53.98	50.29	6
36	31.08	6.26	22.18	21.96	4
37	19.35	6.68	22.61	23.31	4
38	11.20	3.08	9.90	10.06	2
39	34.40	11.61	41.79	41.79	5
40	35.55	6.48	21.69	21.69	6
41	30.53	9.30	35.50	35.14	8
42	25.44	7.10	26.81	26.23	3
43	26.66	11.43	41.36	44.63	6
44	39.79	22.49	84.77	76.12	11
45	8.32	3.64	12.92	13.13	2
46	59.78	13.52	48.80	49.69	15
47	39.22	10.06	37.00	38.33	4
48	3.24	3.18	13.12	12.71	5
49	7.14	5.29	23.10	19.06	8

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

TABLE 3

Unadjusted NFT *Output Observations*

Site #	Total Reading Scores Y_1	Total Math Scores Y_2	Total Coopersmith Scores Y_3
50	39.07	42.71	27.67
51	9.96	14.34	9.33
52	45.37	51.38	31.61
53	18.23	22.05	17.56
54	59.63	64.41	35.89
55	24.20	28.21	18.74
56	13.53	17.09	15.61
57	28.39	27.65	20.79
58	21.67	26.22	13.66
59	120.17	144.67	88.59
60	15.15	18.04	13.58
61	6.92	7.10	6.35
62	9.35	9.85	7.70
63	13.03	13.40	10.29
64	18.63	24.48	23.13
65	12.28	13.01	9.89
66	16.81	19.72	18.70
67	26.36	28.22	24.46
68	22.85	26.21	28.14
69	8.17	8.70	5.12
70	13.69	14.19	12.99

TABLE 4

Unadjusted NFT *Input Observations*

Site #	Education Level of Mother X_1	Occupation Index X_2	Parental Visit Index X_3	Counseling Index X_4	Number of Teachers X_5
50	68.16	12.28	33.58	34.64	15
51	11.88	3.59	13.41	13.82	8
52	55.30	11.53	36.73	35.78	6
53	16.20	7.02	26.94	26.30	9
54	82.45	15.52	45.00	44.23	13
55	15.81	6.93	23.91	23.61	7
56	4.65	5.50	20.91	23.39	5
57	41.25	8.41	26.23	25.24	10
58	10.44	5.22	17.10	18.93	3
59	139.65	35.03	119.56	130.83	22
60	16.28	4.81	18.20	18.98	5
61	12.06	2.59	8.74	8.17	5
62	4.20	2.64	9.89	11.25	2
63	19.44	3.83	12.87	13.23	5
64	28.38	8.91	30.95	33.33	8
65	13.50	3.61	15.60	12.39	4
66	23.32	7.10	24.96	28.56	22
67	27.60	9.38	32.29	34.01	20
68	11.70	10.53	37.67	43.60	8
69	4.68	1.85	6.22	5.46	5
70	10.44	4.82	17.13	18.21	9

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

TABLE 5

PFT *and* NFT *Program Specific α-Envelope Efficiency Values*

PFT Site #	h_0^{*1} Efficiency Value	NFT Site #	h_0^{*2} Efficiency Value
1*	1.00	50	0.95
2	0.90	51	0.92
3	0.98	52*	1.00
4	0.90	53	0.87
5*	1.00	54*	1.00
6	0.90	55*	1.00
7	0.89	56*	1.00
8	0.91	57	0.92
9	0.87	58*	1.00
10*	1.00	59	0.92
11	0.98	60	0.98
12	0.97	61	0.88
13	0.86	62*	1.00
14	0.98	63	0.96
15*	1.00	64	0.91
16	0.95	65	0.97
17*	1.00	66	0.92
18*	1.00	67	0.92
19	0.95	68*	1.00
20*	1.00	69*	1.00
21*	1.00	70	0.94
22*	1.00		
23	0.96		
24*	1.00		
25	0.97		
26	0.93		
27*	1.00		
28	0.94		
29	0.84		
30	0.90		
31	0.83		
32	0.90		
33	0.94		
34	0.85		
35*	1.00		
36	0.80		
37	0.94		
38	0.94		
39	0.91		
40*	1.00		
41	0.94		
42	0.94		
43	0.87		
44*	1.00		
45	0.89		
46	0.90		
47*	1.00		
48*	1.00		
49*	1.00		

*Denotes a site with an efficiency value of "1"

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

desired outputs and designated inputs, as well as the ways in which they are to be measured. We have dealt elsewhere with issues such as efficiency measurement in the presence of inputs that are not subject to managerial discretion,[36] but it would unduly complicate matters to treat that topic here. Hence we shall proceed in the present paper as though the outputs and inputs selected from the Follow Through Project set are all variable, at least for purposes of illustrating the DEA procedures in the present paper.

Repeatedly applying the model (15) to $\alpha = 1$ (Tables 1 and 2) and $\alpha = 2$ (Tables 3 and 4), respectively, we obtain the h_0^{*1} and h_0^{*2} values that are recorded in Table 5. These $h_0^{*\alpha}$ values refer to the efficiency ratings for the DMU's under the PFT and NFT conditions, separately, with $h_0^{*\alpha} = 1$ achievable in either case if, and only if, the DMU being rated is efficient relative to the relevant efficient reference set of DMU's. For example, $h_0^{*1} = 1$ for site 1 means that the DMU at site 1 is rated as efficient relative to some subset of efficient sites in PFT while $h_0^{*2} = 0.94$ for site 70 means that this DMU is only 94% as efficient by reference to the efficient subset of DMU's which are deemed to be relevant among only the NFT set. In each case the relevant comparison set is determined by the model and the computational procedures employed[37] by reference to data in Tables 1 and 2, and Tables 3 and 4 respectively.

Before moving on to the next section, we might once more revert to our earlier comparison with statistical regressions and like approaches. The latter customarily "regress" one output at a time against the inputs to which they are supposedly related.[38] This carries with it very strong assumptions of independence among the indicated outputs[39] in contrast to the methods we use here in which the outputs are all considered *simultaneously* with the indicated inputs to yield the estimates of the extremal relations that underlie the efficiencies shown in Table 5.

6. Evaluating Management Efficiency

As already indicated, it is not readily apparent how one might allow for possibly differing degrees of management efficiency in an experimental design. Something might be done, however, to detect the possible presence of this important source of variation *after* the experimental results are in, and this might be accomplished by statistical methods in association with DEA in various ways.

One possibility would be to compare the relative proportions of $h_0^{*\alpha}$ values which appear in Table 5 for $\alpha = 1$ and $\alpha = 2$. Alternatively (or correlatively) one might also calculate the average of the $h_0^{*\alpha}$ values for both PFT and NFT. One could then test for statistical significance against the null hypothesis of no difference between PFT and NFT by reference to the estimated proportions and averages in order to ascertain whether either set of observations exhibited results that might be attributed to management rather than program efficiency.

Supposing that the results of such tests did not produce statistical significance,[40] one might then proceed to a comparison between the two programs. It is important to realize that such a comparison would continue to be contaminated with managerial inefficiencies—although, of course, this might be justified on grounds that such inefficiencies could not be eliminated, and hence are expected to constitute grounds for part of any choice between different programs.

Approaches to efficiency evaluation via such a DEA-classical statistical testing approach have been conducted in ways like these.[41] We want to go further, however, in order to try to disentangle program efficiency from management efficiency, and reference to the hypothetical situation portrayed in Figure 2 will help us to see what is involved. Here we suppose two sets of DMU's which have similar outputs and inputs that are fixed at the same level for every DMU except for the one input in amounts x and the one output in amounts y shown in Figure 2.

If one were simply to calculate average efficiencies, or even to regress y against x, the set A would rank lower than B. On the other hand, the efficiency frontier for A dominates that of B so that the results of such estimating procedures might lead to erroneous inferences concerning the efficacy of B and A. It would seem

[36] See [24].

[37] E.g., as discussed in [28], use of the simplex or dual methods—or like adjacent extreme point methods of computation—will locate the pertinent set among the basis candidates that these procedures employ. See, however, the corrections also noted in [28] and the subsequent treatment (and proof) in [24].

[38] Such regression approaches were extensively used in the Program Follow Through analyses.

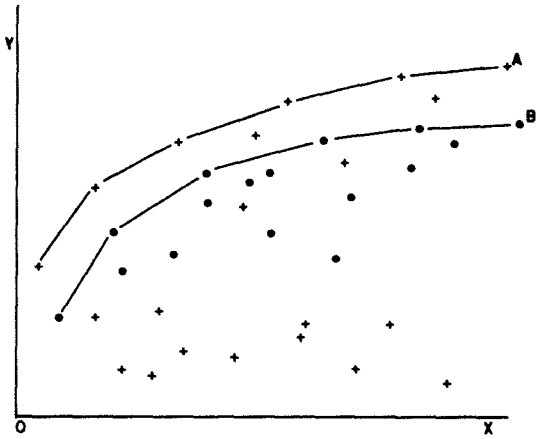
[39] Other approaches which might also have been employed involve simultaneous (econometric) estimation techniques such as are used in [10] and [11].

[40] This was the result secured (i.e., failure to achieve significance) as discussed in [26] and [50].

[41] See, e.g., [26], [50] and [18].

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK



Legend:

Pluses(+) = Observations for DMU's in A
Dots(●) = Observations for DMU's in B

FIGURE 2

prudent, therefore, to at least try to detect the presence of different sources of inefficiency before assigning them all to the programs associated with *A* and *B*, respectively.

We shall shortly be adjusting all points up to the frontiers to effect our comparisons and, of course, the kinds of predictions that one might make from this quarter are different than those one might make by effecting ordinary statistical estimates from the data of Figure 2. In particular, one will now need to effect supplementary analyses and perhaps provide guides and/or controls for the managers associated with the DMU's in such sets as *A* and *B* in order to reenforce the predictions that are being made under our DEA approaches.

The situation is entirely analogous to what one encounters in the standard cost systems that are commonly used to control and guide operations in industrial practice. Under such systems, studies are usually conducted to establish efficient procedures and combinations of material, labor skills, etc., for use in producing specified outputs. So called efficiency variances then signal a departure from one or more of these prescribed approaches and hence provide a basis for inquiry.[42]

The latter result differs, it may be noted, from what would be secured by simply regressing on observations from records of past performance without the intervening engineering studies and controls. To distinguish the latter, we may refer to it as a "control prediction" as distinct from the "pure prediction" that one secures from a regression approach applied to the original (unadjusted) observations. Such a pure prediction assumes that the mixture of efficient and inefficient procedures, material usages, etc. reflected in the observations will continue into the future. The appearance of a variance in the case of a control prediction, on the other hand, indicates the presence of a departure from one or more of the procedures that have been prescribed. In the case of a standard cost system, the predictions for the future continue to be referred back to the coefficients derived from the engineering studies. In our case, these coefficients do not derive solely from the DMU being evaluated. They are obtained from the efficient facet from which the evaluation of this DMU is secured. This means that the values of the efficiency transforms represented by the coefficients of these extremal relations are not known until the efficient facet for the indicated DMU is determined.[43] Once again, therefore, a difference between "control predictions" and "pure predictions" is

[42]Note that such variances are also something more than an index (indicator) number. They often supply measures of the *amount* of inefficiency caused by the departures from standard procedures, materials, etc. (It is of interest to note that Rajiv Banker has now shown how to use these DEA-Efficiency approaches to extend the usual standard-cost variance analyses to the case of joint products. See Rajiv Banker, Studies in Cost Allocation and Efficiency Analyses, DBA Thesis, Boston, Harvard University Graduate School of Business, 1980.)

[43]Mathematical procedures for doing this have been supplied in [28] and [24]. For an illustrative example and discussion, see [9].

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

apparent since the latter, e.g., as secured from the usual regression estimation procedures, are generally applicable to the data from which they were derived without modification, whereas our determinations involve modifications of the data from each DMU by reference to its efficient facet.

Of course, it is theoretically possible for such regression results to coincide with those of DEA as when, for instance, conditions of competition in freely functioning markets produce a state in which all DMU's are at, or very near, the efficiency frontier.[44] We can relate this situation to the one we are considering by regrouping the usual economic conditions for the existence of such possibilities into the following 2 types:

1. *Pressure conditions*: All firms (DMU's) are forced to become as efficient as the most efficient members of the reference set.

2. *Incentive conditions*: The most efficient firms (DMU's) will move to the frontiers that technology makes possible.

The pressure conditions may also be regarded as the condition for survival (into the long run) for any firm under the usual assumptions of market economics. The validity of these assumptions can be tested, as far as the pressure condition is concerned, by reference to data on relative firm performance, even if only in an *ex post facto* manner, by means of measures like profit, costs, etc. A test for the incentive condition, however, is not so readily apparent.

Sans condition two, the pressure conditions allow only for measures of *relative* efficiency. That is, the efficiency of any DMU can be measured only against the members of the set which are characterized as being efficient. Apart from being given access to pertinent information of an *ex cathedra* (e.g., engineering) variety, no means appears to be available for ascertaining whether this efficient subset has reached the frontiers that available technologies and/or programs allow. Within these limits, the procedures we shall now develop may be regarded as an alternate way of producing movement to these *relative* boundaries, or at least evaluating the efficiency losses which may attend any failure to reach them.

7. The Inter Envelope and Program Efficiency

Reference to Figure 3 may help to show how we propose to effect our "program efficiency" comparisons. In this diagram the dots ($\cdot$) and the x's are supposed to represent PFT and NFT observations, respectively. These are all hypothetical data intended to show the amounts of two inputs required to produce one unit of the same output by each of several different DMU's in PFT and NFT, respectively. See the discussion of Figure 1 in section 2.

The observations for PFT and NFT are used to derive the "unit isoquants"[45] that correspond to the α-envelopes for PFT and NFT, respectively. These envelopes and thus the corresponding isoquants are determined by applying the linear programming equivalent of (15) to the observations for $\alpha = 1$, and $\alpha = 2$ in turn. Points such as A represent PFT observations which have values $h_0^{*1} < 1$ since these are not on the PFT α-envelope for which the values $h_0^{*1} = 1$ apply. Similarly, points such as B and C have values of $h_0^{*2} < 1$ determined relative to the NFT-envelope.

Carrying out the repeated applications of (15) needed to determine these envelopes we may find, as in Figure 3, that the PFT α-envelope is more efficient over some regions, and NFT is more efficient over other ranges. Location of such regions may be useful when, for instance, resources may be allocated to both programs in ways which take advantage of these differing efficiencies. In other cases, such as a decision whether to continue PFT, it may be necessary to decide which program is better in some overall efficiency sense.

Such a choice places us in a different position than before. Up to now all of our evaluations and adjustments have involved only improvement in efficiency. A choice between PFT and NFT programs, such as we have just indicated, however, may involve worsening the performance of some DMU's in order to improve the performance of others.

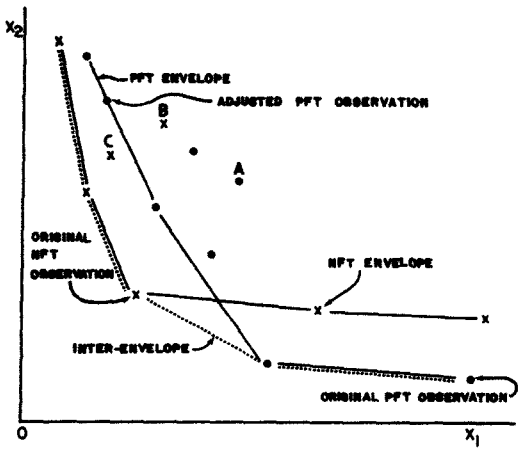
To provide a frame of reference that will allow us to deal with this situation in a knowledgeable manner, we now arrange a new set of hypothetical possibilities. These are based on an adjustment of the original

[44] We are abstracting from statistical error problems such as might be dealt with via the estimation techniques discussed in Bassett and Koenker [8]. Even the latter, however, would require substantial extensions to deal with the multiple outputs and estimation of the piecewise linear relations which we are utilizing in this paper.

[45] See the discussion in §2.

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK



Legend:

● = PFT Observations
× = NFT Observations
●—● = α Envelope for PFT
×—× = α Envelope for NFT
------ = Inter-envelope

FIGURE 3.

observations that will admit any combination of resource allocations and output attainments that either or both programs will allow. This is done as follows:

First we utilize the procedures described in [28] to bring all of the observations onto their respective α-envelopes. Then we construct an inter-envelope that will enable us to compare the resulting clusters of DMU's for each of $\alpha = 1$ and $\alpha = 2$ on the assumption that they are all operating on the efficiency boundaries permitted by their program constraints. We then impute any remaining differences to the respective programs by reference to a common envelope which is always at least as efficient as any of the α-envelopes. Witness, e.g., the "inter-envelope" portrayed in Figure 3.[46]

To make the above more concrete, we now replace (15) with the following formulation for effecting our inter-envelope efficiency determinations.

$$\begin{aligned}
&\max h_0 = \frac{\sum_{r=1}^{s} u_r \hat{y}_{r0}}{\sum_{i=1}^{m} v_i \hat{x}_{i0}} \\
&\text{subject to } 1 \geq \frac{\sum_{r=1}^{s} u_r \hat{y}_{rj}^1}{\sum_{i=1}^{m} v_i \hat{x}_{ij}^1}; \qquad j = 1, \ldots, n_1 \\
&\text{and } 1 \geq \frac{\sum_{r=1}^{s} u_r \hat{y}_{rj}^2}{\sum_{i=1}^{m} v_i \hat{x}_{ij}^2}; \qquad j = 1, \ldots, n_2. \\
&u_r, v_i > 0, \quad \forall i, r.
\end{aligned} \tag{16}$$

[46]For ease of understanding, we are conducting this portion of the discussion as though the isoquant assumption is valid. However, we are dealing with multiple outputs and inputs so that the concept of an isoquant has no meaning and must be replaced by the more general concept of a production possibility set. See [28]. In this same spirit, we shall continue as though we are concerned with resource conservation possibilities. Actually, many of our inputs are fixed, or outside the realm of managerial discretion, as we have noted earlier, and so one might more properly speak of output augmentations rather than input reductions along the lines that are indicated in [27] and [24].

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

Here the caret over a letter indicates that the efficiency adjustments needed to move the DMU's onto the envelopes for $\alpha = 1$ and $\alpha = 2$, respectively, have been carried out as described in the preceding paragraph.[47]

The DMU_0 being rated in (16) can come from either $\alpha = 1$, or $\alpha = 2$, but the h_0^* value is established relative to all DMU's in both programs. Thus, the fact that a DMU is efficient under its program and hence is on its α-envelope, does not necessarily produce an $h_0^* = 1$ for it in this new evaluation. Failure to achieve this rating, however, is now assumed to be due to the program constraint under which this DMU was operating when the adjustments from y_{r0} and x_{i0} to $\hat{y}_{r0}$ and $\hat{x}_{i0}$ were effected. In other words, an $h_0^* < 1$ is now attributed to the program rather than to the efficiency of the management in the DMU being rated.

The indicated application of (16) yields the h_0^* values shown in Table 6. To effect the wanted program evaluation now requires a comparison between the resulting distributions.

One possibility would involve comparisons by some measure—or weighted measure—of the distance between the two distributions. More generally this would need to be extended to a measure of relative distance from some specified reference frame such as the inter-envelope construction that we have just described.[48] In special situations one might also have recourse to considerations of stochastic dominance[49] or even simple inspection. It should suffice for present purposes therefore to note that all of these approaches, including simple inspection of Table 6, lead to the same result:[50] Our evidence is to the effect that PFT has not demonstrated its superior efficiency on the basis of the data we have examined.[51] Thus for reasons advanced earlier (e.g., the additional expenditures involved) an implementation of PFT relative to NFT is not warranted on the basis of our evidence on this portion of the total program.

Of course, our PFT-NFT evaluations need not (and should not) end here. A variety of additional possibilities are also open for study. We might, for instance, conduct a facet-by-facet comparison of the DMU's on each α-envelope.[52] Note, for example, that DMU's on the same facet may be identified explicitly by the fact that they will have the same optimal bases. Furthermore, the direction numbers (and hence the direction cosines)[53] for determining the distance of these DMU's from the relevant part of the inter-envelope can be readily secured and applied facet by facet, if desired, for further evaluation of subsets of PFT-NFT possibilities.

Our construction of a hypothetical reference surface in the form of an inter-envelope opens still further possibilities as well. In particular, it suggests the possibility of a new program involving PFT-NFT combinations. In Figure 3, for instance, the hypothesized situation is such that NFT might be allowed to prevail in the left-hand section of the diagram and PFT in the right-hand section.[54] The broken line segment of the inter-envelope which lies below both α-envelopes indicates the still further possibility of new combinations of PFT and NFT which are more efficient than either of them separately. Such new possibilities need to be confirmed by further study, preferably in the field, but the point is that such possibilities should not be discarded simply because the original experimental design did not consider these potential combinations of PFT and NFT explicitly.

8. Summary and Conclusions

The points that have just been made should suffice to indicate some of the possibilities that our DEA approach may offer. Here we have presented this approach in terms of an illustrative application to Program Follow Through. It is not to be regarded as limited to this Program, however, or even to education programs.

[47] An alternative version is also available which achieves efficiency evaluation as in (16) without first bringing the observations onto their respective α-envelopes. See [50].

[48] The "divergence statistic" of S. Kullback [48] was used for this purpose in an earlier version of this paper [26] that is still available in limited supply from any of the authors. An even more detailed treatment of these data including the development of a suggested canonical form for the statistical distributions may be found in the Appendix to [18].

[49] R. Morey of Duke University reports that he has conducted such an analysis which showed that NFT stochastically dominates PFT.

[50] See preceding two footnotes.

[51] This also seems to conform to the studies by Abt Associates which are based on more complete data as reported in [1] and [2].

[52] See e.g., Gray and Weldon [41]. See also [51] and [14].

[53] See Appendix A in [17] which shows how to obtain the distance from a point to a hyperplane by means of these values.

[54] The information for identifying such facets is available as a byproduct of the computational routines employed.

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

TABLE 6

Inter-Envelope Efficiency Values

PFT Site #	h_0^* Efficiency Value	NFT Site #	h_0^* Efficiency Value
1	0.92	50*	1.00
2*	1.00	51*	1.00
3	0.94	52*	1.00
4*	1.00	53*	1.00
5	0.93	54*	1.00
6*	1.00	55	0.99
7	0.99	56*	1.00
8*	1.00	57*	1.00
9	0.98	58*	1.00
10	0.92	59*	1.00
11*	1.00	60	1.00
12*	1.00	61*	1.00
13	0.99	62*	1.00
14	0.95	63*	1.00
15*	1.00	64*	1.00
16*	1.00	65*	1.00
17*	1.00	66*	1.00
18*	1.00	67*	1.00
19	0.99	68	0.99
20*	1.00	69*	1.00
21*	1.00	70*	1.00
22*	1.00		
23	0.99		
24*	1.00		
25*	1.00		
26	0.99		
27*	1.00		
28*	1.00		
29	0.99		
30*	1.00		
31	0.99		
32*	1.00		
33	0.99		
34	0.98		
35*	1.00		
36*	1.00		
37	0.94		
38	0.99		
39*	1.00		
40	0.95		
41	0.99		
42*	1.00		
43	0.99		
44*	1.00		
45	0.99		
46*	1.00		
47*	1.00		
48*	1.00		
49*	1.00		

*Denotes a site with an efficiency value of "1"

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

Our intention is to provide a general set of concepts and methods that can be applied to a variety of public programs where profit, cost, and like considerations are not directly applicable.[55]

A point to bear in mind, however, is that our DEA approaches and efficiency concepts are at their best when applied to situations in which there is an agreed upon set of objectives and in which resource diversions to other programs are not at issue.[56] Where these conditions are met, there is still an interest in resource conservation, on the presumption that released resources are of use elsewhere. As we have already indicated, our DEA approach gives us a method of ascertaining the *amount* of resource conservation and/or output augmentation involved from improvements in program, or managerial efficiency. How any of the conserved amounts might *best* be redistributed to other activities, e.g., to activities of a non-education variety, involves issues of pricing and weighting that are not addressed in our formulations.

Our remarks at the end of section 6 were directed to showing how the DEA approach differed from customary statistical approaches as well as how the two could be used in various combinations. We might also note that we have here reversed the usual relation between statistical methods and economic theory in empirical research. A great deal of the latter research has been directed toward theory testing, of course, and that is not our objective here. We are instead concerned with using that theory (e.g., accepted parts of production theory) to assist in the evaluation of public policy programs. In addition to arriving at evaluations of these programs (and their management) we have also been concerned with using theory to uncover opportunities for resource conservation or output improvements that would otherwise remain hidden from view.

We have, of course, confined our modeling of possible new opportunities by reference to inferences from observed data. We have also suggested that our DEA approach is best regarded as a guide and that it requires supplementation by further study, preferably in the field, in order to ensure that the indicated opportunities are really present. The fact of their presence is also not decisive unless controls or other alterations can be specified (e.g., by program audits of GAO variety)[57] to ensure that the indicated improvement possibilities will be forthcoming.

In conclusion, we might again contrast our DEA analysis with the "pure prediction" approach that is represented in the following statement by Milton Friedman:[58]

> " . . . The only relevant test of the validity of a hypothesis is comparison of its predictions with experience. . . ."

This is one valid view of the relation between theory and evidence in scientific research, but it is not the only one. It does not emphasize sufficiently the discovery value of a theory as when, for instance, a theory that the earth is round results in a voyage of discovery or when a theory of relativity causes us to search for black holes which have always been there. It also overlooks the possibilities and requirements for control predictions, as when the failure of a mechanism to perform in accordance with its design specifications causes us to correct the mechanism rather than the design. Along the latter route lie opportunities for invention as well as action[59] and these, too, are subjects in need of attention in public programs and management. Their size alone would seem to invite such attention even in those countries which are supposedly capitalist in their orientation. Witness, e.g., Table 7.

In conclusion, we might observe that the opportunities for invention and action may differ greatly in controlled and uncontrolled counterparts.[60] Similarly, the tests used to validate a theory in the latter class of

[55] I.e., costs, profits and other such dollar denominated concepts are not of over-riding importance—but, of course, these concepts can, and probably should, be given a place in the kinds of analyses we are considering by at least considering them among the inputs and outputs to be evaluated. See [28] and [50] for further discussion. See also [6].

[56] In the terminology of [32] we are here concerned with efficiency (including economy) and not "effectiveness" and/or "propriety".

[57] Or by suitably extended versions of such audits to allow for inter-program comparisons and evaluations. See Churchill, *et al.* [32]. For a GAO evaluation of the Program Follow Through study see [56].

[58] From [39, pp. 8–9].

[59] See, e.g., [40] for an example.

[60] See [33] for further discussion.

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

TABLE 7

Total Public Expenditures as Percent of Gross Domestic Product at Market Prices

	1965	1970	1975	1977
Denmark	31	40	46	46
Germany	37	38	48	47
Netherlands	38	44	55	55 (1976)
Norway	34	43	50	51
Sweden	35	43	52	62
United Kingdom	37	41	50	44
United States	27	32	35	33

Sources: Various national statistics and estimates.

Source: T. and F. Geiger, *Welfare and Efficiency* (Washington: National Planning Association, 1978).

cases need not be suited to the situation for control predictions, even when they flow from the same underlying theory.[61]

Appendix

EXHIBIT A

Follow Through Projects, Enrollments and Funding

	1967–68	1968–69	1969–70	1970–71	1971–72	1972–73	1973–74	1974–75	1975–76	1976–77
Number of projects	39	92	160	178	178	173	170	169	165	164
Children enrolled from low income families	2,900	15,500	37,000	60,200	78,170	84,000	81,000	78,000	76,500	75,700
Children rostered in FT			46,650	71,155	90,524	101,974	69,226	89,099	—	—
Rostered Projects			157	177	178	174	172	171	—	—
Federal funds available (in millions of $)	3.75	11.25	32.00	70.30	69.00	63.06	50.62	52.85	55.42	59.00

Source: *The Follow Through Planned Variation Experiment*, Volume V, The Follow Through Evaluation: A Technical History, p. 21 in a report prepared for the U.S. Department of Health, Education and Welfare by W. Haney of the Huron Institute, Cambridge, Ma. 02138 August, 1977.

[61]The authors are grateful to Professor Arie Lewin and several referees whose comments are incorporated in this extensive revision of an earlier manuscript, referenced as [26] in the bibliography, which served as the basis for a presentation at the TIMS/ORSA meetings in New York on May 1, 1978. 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 NR 1947-021, ONR Contract N00014-75-C-0616 with the Center for Cybernetic Studies, The University of Texas, and ONR Contract N0014-76-C-0932 at Carnegie-Mellon University's School of Urban and Public Affairs. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

EXHIBIT B

Follow Through Approaches and Associated Sponsors Included in Data Envelopment Analysis Study

Approach and Sponsor	Number of PFT Sites	Number of NFT Sites
RESPONSIVE EDUCATION PROGRAM Far West Laboratory for Educational Research and Development	7	7
TUCSON EARLY EDUCATION MODEL (TEEM) Arizona Center for Early Childhood Education	4	2
BANK STREET COLLEGE OF EDUCATION APPROACH Bank Street College of Education	5	3
DIRECT INSTRUCTION MODEL (DIM) University of Oregon–College of Education	5	3
BEHAVIOR ANALYSIS APPROACH (BA) Support and Development Center for Follow Through—University of Kansas	6	1
COGNITIVELY ORIENTED CURRICULUM MODEL High/Scope Educational Research Foundation	3	—
FLORIDA PARENT EDUCATION MODEL University of Florida	4	3
EDC OPEN EDUCATION FOLLOW THROUGH PROGRAM Education Development Center	2	—
SELF-SPONSORED—New York City, NY	1	—
SELF-SPONSORED—Philadelphia, PA	1	1
SELF-SPONSORED—Detroit, MICH	1	—
SELF-SPONSORED—Portland, OR	1	—
SELF-SPONSORED—San Diego, CA	1	—
INTERDEPENDENT LEARNING MODEL (ILM) New York University—Institute for Developmental Studies	1	—
LANGUAGE DEVELOPMENT (BILINGUAL) EDUCATION APPROACH Southwest Educational Development Laboratory (SEDL)	2	1
HOME-SCHOOL PARTNERSHIP: A MOTIVATIONAL APPROACH Southern University and A&M College	1	—
CALIFORNIA PROCESS MODEL California State Department of Education–Division of Compensatory Education	4	—
Total Number of Sites	49	21

Source: Abt [2], p. A-18

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

EXHIBIT C

Site Level Distribution of DEA Study Sample

PFT Site #	NFT Site #	Model and Site Name	Region[1]	City Size[2]	PFT[3] Student Pop.	NFT[4] Student Pop.
		Responsive Education Model				
1	50	Berkeley, CA	W	Medium City	99	71
2	51	Buffalo, NY	NE	Large City	77	27
3	52	Duluth, MN	NC	Medium City	77	79
4	53	Fresno, CA	W	Medium City	48	54
5	54	Lebanon, NH	NE	Rural Area	14	97
6	55	Salt Lake, UT	W	Medium City	36	51
7	56	Tacoma, WA	W	Medium City	51	42
		TEEM Model				
8		Baltimore, MD	S	Large City	99	—
9		Lakewood, NJ	NE	Small City	80	—
10	57	Lincoln, NB	NC	Medium City	96	55
11	58	Wichita, KS	NC	Large City	84	36
		Bank Street Model				
12	59	New York, NY	NE	Large City	72	245*
13	60	Philadelphia, PA	NE	Large City	80	37
14		Brattleboro, VT	NE	Small City	20	—
15	61	Fall River, MA	NE	Medium City	39	18
16		Wilmington, DE	S	Medium City	109	—
		DIM Model				
17		New York, NY	NE	Large City	31	—
18	62	E. St. Louis, IL	NC	Large City	56	21
19		Grand Rapids, MI	NC	Medium City	103	—
20	63	Racine, WI	NC	Medium City	62	27
21	64	Flint, MI	NC	Medium City	77	66
		BA Model				
22		New York, NY	NE	Large City	43	—
23	65	Philadelphia, PA	NE	Large City	108	27
24		Portageville, MO	NC	Rural Area	47	—
25		Kansas City, MO	NC	Large City	61	—
26		Louisville, KY	S	Large City	90	—
27		Meridian, IL	NC	Rural Area	68	—
		Cognitive Curriculum Model				
28		New York, NY	NE	Large City	52	—
29		Chicago, IL	NC	Large City	18	—
30		Okaloosa Co., FL	S	Small City	48	—
		Parent Education Model				
31		Philadelphia, PA	NE	Large City	46	—
32	66	Jacksonville, FL	S	Large City	15	53
33	67	Richmond, VA	S	Large City	111	69
34	68	Houston, TX	S	Large City	95	78
		EDC Model				
35		Philadelphia, PA	NE	Large City	112	—
36		Paterson, NJ	NE	Medium City	42	—

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

EXHIBIT C (*continued*)

PFT Site #	NFT Site #	Model and Site Name	Region[1]	City Size[2]	PFT[3] Student Pop.	NFT[4] Student Pop.
		Self-Sponsored Model				
37		Detroit, MI	NC	Large City	43	—
38	69	New York, NY	NE	Large City	20	13
39		Philadelphia, PA	NE	Large City	86	—
40		Portland, OR	W	Large City	45	—
41		San Diego, CA	W	Large City	71	—
		ILM Model				
42		New York, NY	NE	Large City	53	—
		SEDL Model				
43	70	Philadelphia, PA	NE	Large City	86	36
44		Tulare, CA	W	Small City	173	—
		Home-School Partnership Model				
45		New York, NY	NE	Large City	26	—
		California Process Model				
46		Los Angeles, CA	W	Large City	98	—
47		Ravenswood, CA	W	Small City	74	—
48		Lamont, CA	W	Rural Area	27	—
49		San Jose, CA	W	Large City	42	—
				Total Student Pop.	3210	1202

*Pooled Citywide NFT Population Figure for New York City

[1]NE = North Eastern United States
S = Southern United States
NC = North Central United States
W = Western United States

[2]Large City = 200,000 or more
Medium City = 50,000 to 199,999
Small City = 10,000 to 49,999
Rural Area = Less than 10,000

[3]All Data Envelopment Analysis study information refers to the Cohort II-K student population. II-K indicates that this group of students began their Program Follow Through experience in kindergarten. (This was also the only one of three Cohorts which had completed all of the grades from kindergarten through third grade at the time of our study for which site level information was available). However, due to incomplete statistics along some DEA variable dimensions, some of the Cohort II-K PFT sites were not included in the DEA study. Specifically, Bank Street Model: Rochester, NJ site; EDC Model: Chicago, IL site; and SEDL Model: St. Martin Parish, LA site were excluded from the DEA study student population. The actual Cohort II-K PFT population was 3,367 of which, as noted above, a set of 3,210 students were used in the DEA study. This exclusion of sites also extended to the NFT groups which were similarly reduced to 1,202 students.

[4]Two sets of NFT student groups were created in the original Program Follow Through study. One group was a local student set, usually in the same school system as the subject PFT site. The second group, and the one selected for the DEA study, was a "best matched" group, which may or may not have been located in the same school system or even the same geographical region. The NFT group which most nearly matched the PFT students of a given site along a number of demographic and initial performance dimensions was considered the "best match" for the latter. For several PFT sites the *same* "best matched" NFT group was used. The much smaller NFT student population total of 1,202 as compared to the PFT student total of 3,210 resulted. See also preceding footnote.

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

References

1. ABT ASSOCIATES, *Education as Experimentation: A Planned Variation Model*, Vol. IA and IB, Abt Associates, Inc., Cambridge, Mass., 1974.
2. ———, *Education as Experimentation: A Planned Variation Model*, Vol. IIIA and IIIB (Cambridge, Mass: Abt Associates, Inc. 1976). Also issued as *The Follow Through Planned Variation Experiment*, Vols. IIA, IIB and IIC, U.S. Office of Education, Washington, 1977.
3. AIGNER, D. J., AMEMIYA, T. AND POIRIER, P. J., "On the Estimation of Production Frontiers: Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function," *Internat. Econom. Rev.*, Vol. 17, No. 2 (1976), pp. 377–396.
4. AKAIKE, A., "Information Theory and an Extension of the Maximum Likelihood Principle," *Proceedings of the Second International Symposium of Information Theory*, Hungarian Academy of Sciences, Budapest, 1972.
5. ALLEN, R. G. D., *Mathematical Analysis for Economists*, MacMillan, New York, 1939.
6. BANKER, R., "A Game Theoretic Approach to Measuring Efficiency," *European J. Operational Res.* (to appear).
7. ———, CHARNES, A., COOPER, W. W. AND SCHINNAR, A., "A Bi-Extremal Principle for Estimating the Efficiency of Decision Making Units," *Management Sci.* (submitted).
8. BASSETT, G. AND KOENKER, R., "Estimation of Efficiency Contours in Flexible Models of Production," A paper prepared for presentation at the 1979 winter meetings of the Econometric Society in Atlanta, Bell Telephone Laboratories, Inc., Murry Hill, N.J.
9. BESSENT, A. AND BESSENT, W., "Determining Attributes of Efficient and Inefficient Schools Through Data Envelopment Analysis," *Educational Admin. Quart.* (to appear).
10. BOARDMAN, A., "Policy Models for the Management of Student Achievement and Other Education Outputs," *TIMS Studies in the Management Sciences*, Vol. 8, Management Science Approaches to Manpower Planning and Organization Design, North-Holland, Amsterdam, 1978.
11. ———, DAVIS, O. A. AND SANDAY, P. R., "A Simultaneous Equations Model of the Educational Process," *J. Public Econom.*, Vol. 7 (1977), pp. 23–49.
12. ——— AND MURNANE, R. J., "The Use of Panel Data in Education Research," *Sociology of Education* (to appear).
13. CAMPBELL, D. T. AND STANLEY, J. C., *Experimental and Quasi-Experimental Designs for Research*, Rand-McNally, Chicago, Ill., 1963.
14. CARLSON, D. E., *The Production and Cost Behavior of Higher Education Institutions*, University of California, Berkeley, Calif., 1972.
15. CARLSON, S., *The Pure Theory of Production*, P. S. King, London, 1939.
16. CHARNES, A. AND COOPER, W. W., "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.
17. ——— AND ———, *Management Models and Industrial Application of Linear Programming*, Vol. 1, Wiley, New York, 1961.
18. ——— AND ———, "Management Science Relations for Evaluation and Management Accountability," *J. Enterprise Management*, Vol. 2, No. 2 (1980), pp. 143–162.
19. ——— AND ———, "Managerial Economics—Past, Present and Future," *J. Enterprise Management*, Vol. I, No. 1 (1978), pp. 5–23.
20. ———, ———, DEVOE, J. K. AND LEARNER, D. B., "Demon, Mark II: An Extremal Equation Approach to New Product Marketing," *Management Sci.*, Vol. 14, No. 9 (1968), pp. 513–524.
21. ———, ———, ——— AND ———, "Demon, Mark II: Extremal Equations Solution and Approximation," *Management Sci.*, Vol. 14, No. 11 (1968), pp. 682–691.
22. ———, ——— AND LEARNER, D. B., "Constrained Information Theoretic Characterizations in Consumer Purchase Behavior," *J. Operational Res. Soc.*, Vol. 29, No. 9 (1978), pp. 833–884.
23. ———, ———, ——— AND PHILLIPS, F., "The MDI Method as a Generalization of Logit, Probit and Hendry Analyses in Marketing," A paper presented at the 1980 TIMS/ORSA conference in Washington, D.C., Market Research Corporation of America, Stanford, Conn., May 1980.
24. ———, ———, 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, October 1980.
25. ———, ——— AND MELLON, B., "Blending Aviation Gasoline—A Study in Programming Interdepen-

RIGHTSLINK

dent Activities in an Integrated Oil Company," *Econometrica*, Vol. 20, No. 2 (1952). See also September, 1952, for an exchange between A. Manne and A. Charnes.

26. ———, ——— AND RHODES, E., "A Data Envelopment Analysis Approach to Evaluation of the Program Follow Through Experiment in U.S. Public School Education," Management Sciences Research Report No. 432, Carnegie-Mellon University School of Urban and Public Affairs, Pittsburgh, November, 1978. Subsequently revised and issued in May, 1980 as "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".
27. ———, ——— AND ———, "An Efficiency Opening for Managerial Accounting in Not-For-Profit Entities," in P. Holzer, Editor, *Proceedings of a Conference on Managerial Accounting*, University of Illinois Department of Accountancy, Urbana, 1980.
28. ———, ———, AND ———, "Measuring the Efficiency of Decision-Making Units," *European J. Operational Res.*, Vol. 2, No. 6 (1978), pp. 429–444. See also "Corrections," *op.cit.*, Vol. 3, No. 4 (1979), p. 339.
29. ———, ——— AND ———, "On the Distribution of Efficiency Measures for Decision Making Units," University of Texas Center for Cybernetic Studies, Research Memorandum, Austin, Texas, Feb. 1978. Included as an Appendix in [18].
30. ———, ——— AND SCHINNAR, A., "Transforms and Approximations in Cost and Production Function Relations," Research Report CCS 339, The University of Texas at Austin, Center for Cybernetic Studies, January, 1977.
31. ———, ——— AND SEIFORD, L., "Extremal Principles and Optimization Dualities for Khinchin-Kullback-Leibler Estimation," *Math. Operationsforsch. Statist. Ser. Optimization*, Vol. 9, No. 1 (1978), pp. 21–29.
32. CHURCHILL, N. C., COOPER, W. W., SAN MIGUEL, J., GOVINDARAJAN, V. AND POND, J., "Comprehensive Audits: Some Findings and Some Suggestions for Research," *Symposium on Auditing Research*, II, Department of Accounting, University of Illinois at Urbana-Champaign, 1977.
33. COOPER, W. W., "Understanding Prediction and Control—And Other Matters Relating to Scientific Research," *SUPALUM* Alumni Magazine, School of Urban and Public Affairs, Carnegie-Mellon University, No. 2, April, 1974, pp. 25–28.
34. COOPERSMITH, S., *The Antecedents of Self-Esteem*, San Francisco, Calif.
35. FÄRE, R. AND LOVELL, C. A. K., "Measuring the Technical Efficiency of Production," *J. Econom. Theory*, Vol. 19 (1978), pp. 150–162.
36. FARRELL, M. J., "The Measurement of Productive Efficiency," *J. Roy. Statist. Soc.*, Vol. 120, Sec. A, pt. 3 (1957), pp. 253–290.
37. FERBER, R. AND HIRSCH, W. Z., "Social Experiments in Economics," in D. J. Aigner and C. N. Morris, eds., "Experimental Design in Econometrics," Annals of Applied Econometrics 1979-2, Supplement to *J. Econometrics*, Vol. 11, No. 1 (1979).
38. FØRSUND, F. R., KNOX LOWELL, C. A. AND SCHMIDT, P., "A Survey of Frontier Productions and of Their Relationship to Efficiency Measurement," *J. Econometrics* (May 1980).
39. FRIEDMAN, M., *Essays in Positive Economics*, University of Chicago Press, Chicago, Ill., 1953.
40. FUMAS, V. S. AND WHINSTON, A. B., "Flexible Contracting Theory and Case Examples," *European J. Operational Res.*, Vol. 3, No. 5 (1979), pp. 368–378.
41. GRAY, R. AND WELDON, K., "An Experiment with Convex Production Functions," Working Paper. National Center for Higher Education Management Systems, Boulder, Colo., Nov. 1978.
42. GRILICHES, Z., "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, Vol. 14, No. 1 (1977), pp. 1–22.
43. HANEY, W., "The Follow Through Planned Variation Experiment," *The Follow Through Evaluation: A Technical History*, Vol. 5. Prepared for the Office of Planning, Budgeting and Evaluation of the U.S. Education Department of HEW by The Huron Institute, Cambridge, Mass., August, 1977.
44. HANUSHEK, E. A., "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *J. Human Resources*, Vol. 14, No. 3 (1979).
45. ———, "Families, Schools and Achievement," Discussion paper 7906, University of Rochester Public Policy Analysis Program, Rochester, N.Y., 1979.
46. HOLT, C. C., MODIGLIANI, F., MUTH, J. F. AND SIMON, H. A., *Planning Production, Inventories and Work Force*, Prentice-Hall, Englewood Cliffs, N.J., 1960.
47. KENNINGTON, J., "A Primal Simplex Code for Computing the Efficiency of Decision Making Units," Technical Report OREM 80001, Dept. of Operations Research, Dallas: Southern Methodist University, January, 1980.

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

48. KULLBACK, S., *Information Theory and Statistics*, Wiley, New York, 1959.
49. *Metropolitan Achievement Tests Special Report*, Harcourt Brace Jovanovich, New York, 1971.
50. RHODES, E. L., *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*, unpublished Ph.D. thesis, Carnegie-Mellon University School of Urban and Public Affairs, Pittsburgh, Pa., 1978.
51. RODGERS, K. W., "The Realization of National Policy Objectives by Historically Black Colleges," Arthur D. Little, Cambridge, Mass., 1976.
52. SAWA, T., "Information Criteria for Discriminating Among Alternative Regression Models," *Econometrica*, Vol. 46, No. 6 (1978), pp. 1273–1291.
53. STALLINGS, JANE A., "Follow Through Program Classroom Observation Evaluation, 1971–72," Stanford Research Institute, Menlo Park, Calif., 1973.
54. SUMMERS, A. A. AND WOLFE, B. L., "Which School Resources Help Learning? Efficiency and Equity in Philadelphia Public Schools," *Business Review of Federal Reserve Bank of Philadelphia*, Feb., 1975. See also Summers and Wolfe "Do Schools Make a Difference?," *Amer. Econom. Rev.* (to appear).
55. THEIL, H., *Optimal Decision Rules for Government and Industry*, North-Holland, Amsterdam, 1964.
56. U.S. GENERAL ACCOUNTING OFFICE, *Follow Through: Lessons Learned from Its Evaluation and Need to Improve Its Administration*, U.S. General Accounting Office, Washington, 1975.

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK

Copyright 1981, by INFORMS, all rights reserved. Copyright of Management Science is the property of INFORMS: Institute for Operations Research and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Downloaded from informs.org by [130.39.62.90] on 20 August 2014, at 12:11 . For personal use only, all rights reserved.

RIGHTSLINK