

AD-A168 588

DEVELOPMENTS IN EFFICIENCY ANALYSIS

A. Charnes, W. W. Cooper, B. Golany,
E. Schmitz, H. Sherman, and J. Stutz

The University of Texas at Austin

for

BASIC RESEARCH LABORATORY
Milton R. Katz, Director



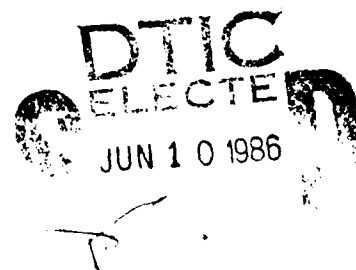
U. S. Army

Research Institute for the Behavioral and Social Sciences

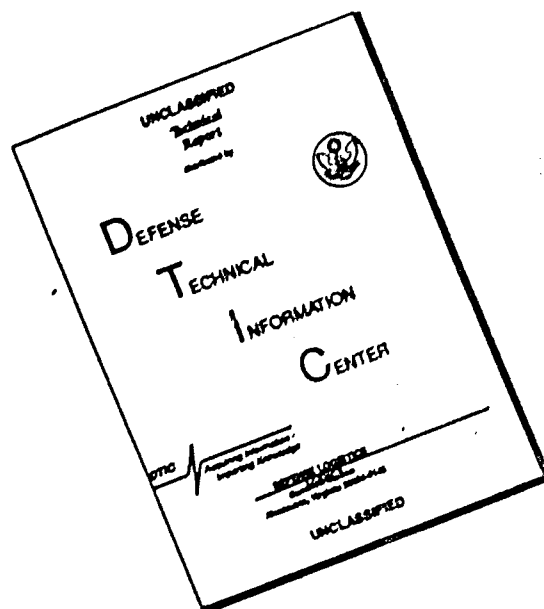
May 1986

Approved for public release; distribution unlimited

DTIC FILE COPY



DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the
Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Technical Review By:

Hyder Lakhani
Abraham Nelson

This report, as submitted by the contractor, has been cleared for release to Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or other reference services such as the National Technical Information Service (NTIS). The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Research Note 86-61	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) DEVELOPMENTS IN EFFICIENCY ANALYSIS		5. TYPE OF REPORT & PERIOD COVERED Period ending August 1984
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) A. Charnes, W.W. Cooper, B. Golany, H. Sherman, and J. Stutz		8. CONTRACT OR GRANT NUMBER(s) MDA 903-81-C-0365
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Cybernetic Studies, The University of Texas at Austin, Austin, Texas 78712		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences, 5001 Eisenhower Avenue, Alexandria, VA 22333-5600		12. REPORT DATE May 1986
		13. NUMBER OF PAGES 77
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) --		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE --
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) --		
18. SUPPLEMENTARY NOTES Edward Schmitz, contracting officer's representative		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Multiplicative Efficiency Pareto Optimality Validation Log Efficiency Efficiency Analysis Data Envelopment Analysis Statistical Regression Piecewise Cobb-Douglas Forms Ratio Analysis Production Functions Simulation		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Concepts of efficiency analysis are extended through the Data Envelopment Analysis method to include new theoretical characterizations of Pareto optimal production functions, solutions of such problems as economies of scale, non-concavity and isotonicity, discretionary and non-discretionary inputs, multiplicative efficiency formulations (piecewise Cobb-Douglas), and comparisons to alternatives such as ratio and regression analysis.		

DEVELOPMENTS IN EFFICIENCY ANALYSIS

EXECUTIVE SUMMARY

Requirement:

The US Army Research Institute conducts research on manpower, personnel, and training issues of interest to the Army. Recently there has been a need for basic research into objective ways of measuring and evaluating organizational performance efficiency under different policies and resource allocation procedures.

Procedure:

Building upon their previous theoretical work in mathematical programming, the authors have generalized the concept of Data Envelopment Analysis to include new theoretical characterizations of empirical production functions.

Findings:

The developments show how a Pareto-Optimal frontier production function can be developed, and such problems as economies of scale, isotonicity and non-concavity, discretionary and non-discretionary inputs, and Cobb-Douglas multiplicative functional problems can be solved. Also, simulations are performed which demonstrate that DEA methodology is not only superior to other methods (ratio analysis and regression analysis) for identifying inefficiencies but also for locating their sources and estimating their magnitude in particular decision making units.

Utilization of Findings:

Methodologies developed here provide new approaches for measuring the efficiency and productivity of organizations that have multiple inputs and outputs. This methodology could be applied to resource allocation and evaluation problems in recruiting, training, unit performance, equipment maintenance, personnel management, logistic management, and weapon system development.

DEVELOPMENTS IN EFFICIENCY ANALYSIS

CONTENTS

	Page
I. INTRODUCTION.....	1
II. PARETO OPTIMALITY, EFFICIENCY ANALYSIS, AND EMPIRICAL PRODUCTION FUNCTIONS.....	4
III. INVARIANT MULTIPLICATIVE EFFICIENCY AND PIECEWISE COBB-DOUGLAS ENVELOPMENTS.....	26
IV. A COMPARATIVE STUDY OF DATA ENVELOPMENT ANALYSIS AND OTHER APPROACHES TO EFFICIENCY EVALUATION AND ESTIMATION.....	31
APPENDIX.....	70

LIST OF TABLES

Table 1.	DEA Ratings of Artificial DMUs.....	44
2.	Single Output Measures.....	47
3.	Comparison of DEA, Ratio Analysis, and Linear Regression Approaches Ability to Locate Inefficient DMUs.....	56
4.	H15 Intensity Adjustment and Efficiency Value....	58

LIST OF FIGURES

Figure 1.	Empirical Production Possibility Set.....	7
2.	Isotonic Function with Concave Cap.....	20
3.	Concave Production Frontier.....	20

I. INTRODUCTION

Economists and management scientists have long been interested in production functions, or the relationship of resources to organizational outputs. Data envelopment analysis (DEA), developed by Charnes, Cooper, and Rhodes (1978) provides a new methodology for measuring the technical efficiency of organizations that use multiple inputs to produce multiple outputs.

Data envelopment analysis has contributed to both basic and applied research in efficiency analysis. It is basic in the sense that it provides a new mathematical model for describing behavior of organizations in the transforming inputs to outputs. It is applied since it relies upon empirical data with direct implications for identifying specific inefficiencies and redirecting management effort. It is ideally suited for the evaluation of public sector institutions, because it can deal with multiple outputs and does not require information on prices. DEA has been applied to education (Bessent, 1983), health care (Sherman, 1981), Navy recruiting (Lewin and Morey, 1980), criminal court systems (Lewin and Morey, 1984), and computer software evaluation (Barr, 1983).

The following sections of this report describe basic research that has extended and improved the mathematical models available for analyzing organizational efficiency. Section II provides a new method of data envelopment analysis methodology that is a substantial improvement over the original approach. This new model permits the analysis of the rates of change of individual outputs with respect to change in specific inputs. Further, the new model improves the computational algorithm by only searching the optimal points in the solution space.

Section III provides a multiplicative efficiency model. Previous formulations had been sensitive to the units of measurement. Here, a simple change is formulated that preserves the desirability of the multiplicative format and creates invariant measures of efficiency.

The last section compares DEA, ratio, and regression analysis through investigation of an artificial data base. The results favor DEA not only for identifying inefficiencies but also locating their sources and amounts. The advantage of DEA is that it performs a separate optimization for each observation and does not attempt to capture a great varieties of behaviors in a smooth and simple functional form.

Efficiency analysis as developed and extended in this report, contains substantial potential for improving the resource allocation and evaluation process within the Army. An application has already been made to recruiting research management (Charnes, 1982). Additional areas that could benefit from efficiency analysis include training unit performance, weapon system development, equipment maintenance, logistic management, and personnel management.

REFERENCES

- Barr, R.S., J.L. Kennington, and P. McKee. "A New Efficiency Approach to Comparative Testing of Mathematical Programming Codes." Paper presented at TIMS/ORSA Joint National Meeting. April 1983.
- Bessent, A., W. Bessent, J. Kennington, and B. Reagan. "An Application of Mathematical Programming to Assess Productivity in the Houston Independent School District." Management Science. Dec 1982.
- Charnes, A., W.W. Cooper, D. Divine, J. Klopp, and J. Stutz. "An Application of Data Envelopment Analysis to Management of Army Recruiting Districts." Paper presented at ORSA/TIMS Joint National Meeting. Oct 1982.
- Charnes, A., W.W. Cooper, and E. Rhodes. "Measuring the Efficiency of Decision-Making Units." European Journal of Operations Research. Nov 1978.
- Lewin, A.Y. and R.C. Morey. "Studies to Validate and Extend the Data Envelopment Analysis Method for Measuring Relative Efficiency of Decision Making Units." Duke Univ., Fuqua School of Business. 1980.
- Lewin, A.Y. and R.C. Morey. "Evaluating the Administrative Efficiency of Courts." Omega. 1983.
- Sherman, H.D. "Measurement of Hospital Technical Efficiency: A Comparative Evaluation of Data Envelopment Analysis and Locating Inefficiency in Health Care Organizations. Unpublished Doctoral Dissertation, Harvard Univ. Graduate School of Business Administration. 1981.

II. PARETO OPTIMALITY, EFFICIENCY ANALYSIS AND EMPIRICAL PRODUCTION FUNCTIONS

Classically, the economic theory of production is heavily based on the conceptual use of the Pareto-efficiency (or Pareto-optimal) frontier of production possibility sets to define "the" production function. The work of R. Shephard [18], [19] under severe restrictions on the mathematical structure of production possibility sets and cost relations, developed an elegant "transform" theory between production aspects and cost aspects [10]. This was applied to various classes of explicitly given parametric functional forms and problems of statistical estimation of parameters from data were considered in classical statistical contexts especially by successors such as R. Frisch, S. Afriat, D. Aigner, F. Forsund [1, 2, 16]. These efforts were almost exclusively for single output functions.

M.J. Farrell in [14], seeking to disentangle prices or costs from "technical" aspects of production, as well as to provide a more meaningful technical setting to statistical and empirical aspects of production, defined (for the single output case) a measure of "technical efficiency" of observed production units relative to the total units observed assuming that the production process of inputs to output conversion was linear and of constant returns to scale.

Building on the unit-by-unit evaluations of Farrell and the engineering ratio idea of efficiency measure for a single input and output, efficiency analysis in its managerial aspects and its constructible extensions to multi-input, multi-output situations was initiated by Charnes, Cooper and Rhodes in [8], [9]. Subsequent extensions and elaborations by the former pair with other students and colleagues were made in [7], [11], [12] . . . with more attention to classical economic aspects and to

the production function side of the mathematical duality structure and Data Envelopment Analysis first discovered in the CCR work. The CCR ratio measures and the variants of Farrell, Shephard, Färe, Banker, et al. require, however, non-Archimedean constructs for rigorous theory and usage. Their solution methods also do not easily provide important needed properties of their associated empirical production functions.

Thus, in this paper we introduce as basic the idea of Pareto optimality with respect to an empirically defined production possibility set. We characterize the mathematical structures permitted under our minimal assumptions and contrast these with others' work. Properties such as isotonicity, non-concavity, economies of scale, piece-wise linearity, Cobb-Douglas forms, discretionary and non-discretionary inputs are treated through a new Data Envelopment Analysis method and informatics which permits a constructive development of an empirical production function and its partial derivatives without loss of efficiency analysis or use of non-Archimedean field extensions.

EMPIRICAL FUNCTION SETTING AND GENERATION

By an "empirical" function we shall mean a vector function whose values are known at a finite number of points and whose values at other points in its domain are given by linear (usually convex) combinations of values at known points. The points in the domain are "inputs," the component values of the vector function "outputs." We shall assume that inputs are so chosen that convex combinations of input values for each input are meaningful input values. We assume this for output values as well.

In efficiency analysis, observations are generated by a finite number of "DMU"s, or "productive," or "response" units, all of which have the same inputs and outputs. A relative efficiency rating is to be obtained for each unit. Typically, observations over time will be made of each unit and the results of efficiency analyses will be employed to assist in managing each of the units. We assume n units, s outputs and m inputs. The values are to be non-negative (sometimes positive) numbers.

A HYPOGRAPH EMPIRICAL PRODUCTION POSSIBILITY SET

Given the (empirical) points (X_j, Y_j) , $j=1, \dots, n$ with $(m \times 1)$ "input" vectors $X_j \geq 0$ and $(s \times 1)$ "output" vectors $Y_j \geq 0$, we define the "empirical production set" P_E to be the convex hull of these points i.e.

$$(2.1) \quad P_E \triangleq \{(x, y) : x = \sum_{j=1}^n X_j \nu_j, y = \sum_{j=1}^n Y_j \nu_j, \forall \nu_j \geq 0, \sum_j \nu_j = 1\}$$

We extend it to our "empirical production possibility set" Q_E by adding to P_E all points with inputs in P_E and outputs not greater than some output in P_E i.e.

$$(2.2) \quad Q_E \triangleq \{(x, y) : x = \bar{x}, y \leq \bar{y} \text{ for some } (\bar{x}, \bar{y}) \in P_E\}$$

Note that Q_E is contained in (e.g. is smaller than) every production possibility set heretofore employed, i.e. those studied by Farrell [14], Shephard [19], Banker, Charnes and Cooper [3], Färe, et al. [13], etc. The Farrell, Shephard, Färe sets are (truncated) cones; the BCC set (when not also a cone) adds to Q_E the set

$$\{(x, y) : x \geq \bar{x}, y = \bar{y} \text{ for some } (\bar{x}, \bar{y}) \in Q_E\}.$$

These relations may be visualized in the schematic plot of

Figure 1:

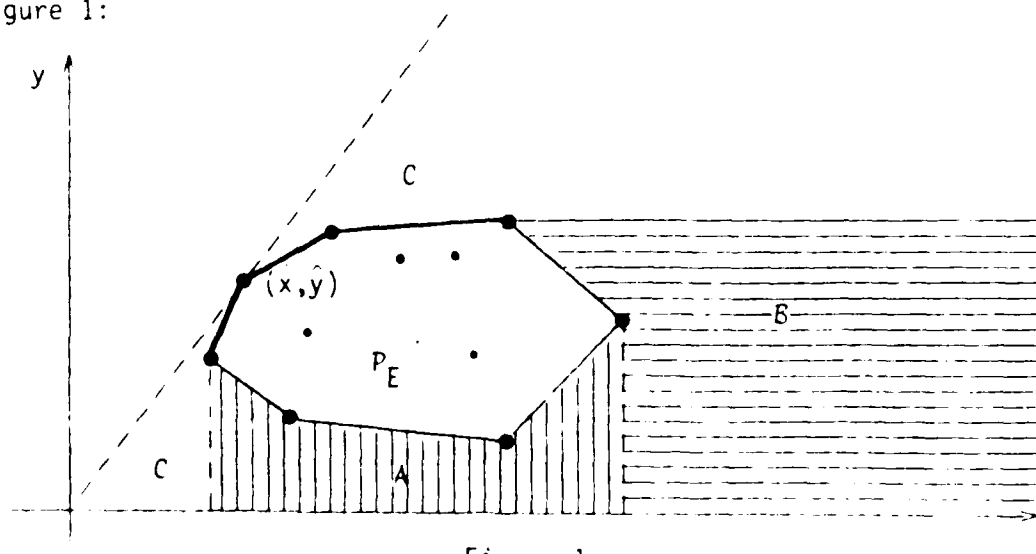


Figure 1

where $Q_E = P_E \cup A$, the BBC set is $Q_E \cup B$, and the Farrell, Shephard, Färe set is $Q_E \cup B \cup C$.

Let P_E^h , Q_E^h denote the sets corresponding to P_E and Q_E when only the output y_h is the ordinate. Evidently a frontier function $f_h(x)$ is determined by

$$(2.3) \quad f_h(x) = \max y_h \text{ for } (x, y_h) \in Q_E^h$$

Then,

Theorem 0: Q_E^h is the hypograph of $f_h(x)$ over $\{x : (x, y) \in Q_E\}$

Proof: The hypograph H_h of $f_h(x)$ is the set

$$H_h \triangleq \{(x, y_h) : y_h \leq f_h(x), (x, y) \in Q_E\}$$

Let $\mathcal{D}_E$ denote $\{x : (x, y) \in Q_E\}$. It is the domain (the input set) of our empirical frontier functions.

Theorem 1: $f_h(x)$ is a concave, piecewise linear function on $\mathcal{D}_E$.

Proof: A necessary and sufficient condition for $f_h(x)$ to be concave is that its hypograph is a convex set (cf. Rockefellar [17], or Fenchel [15]). The piecewise linearity also follows from the construction of Q_E by all convex combinations of the empirical points (X_j, Y_j) , $j=1, \dots, n$.

We observe explicitly further that no use whatever has been made of non-negativity of input and output values in the sets, functions or proof of Theorems 0 and 1. Therefore, they hold without this restriction--a fact we shall employ elsewhere.

Also, no assumptions have been made about the properties of any underlying function, or function hypograph, from which the (X_j, Y_j) of our empirical construct may be considered samples. Theorem 1 shows, therefore, that any empirical (maximum) frontier function is the "concave cap" function of its graph.

THE EMPIRICAL PARETO-OPTIMAL PRODUCTION FUNCTION

A Pareto-optimum for a finite set of functions $g_1(x), \dots, g_K(x)$ is a point x^* such that there is no other point x in the domain of these functions such that

$$(3.1) \quad g_k(x) \leq g_k(x^*) \quad , \quad k=1, \dots, K$$

with at least one strict inequality. Charnes and Cooper in [5], Chapter IX, showed that x^* is Pareto-optimal iff x^* is an optimal solution to the mathematical (goal) program

$$(3.2) \quad \min \sum_{k=1}^K g_k(x) \quad \text{subject to } g_k(x) \leq g_k(x^*) \quad , \quad k=1, \dots, K$$

This was employed by Ben-Israel, Ben-Tal and Charnes in [4] to develop the currently strongest necessary and sufficient conditions for a Pareto-optimum in convex programming.

Utilizing (3.2) we can now define and construct, im(or ex-)plicitly the Pareto-optimal (or "Pareto-efficient") empirical (frontier) production function. Other usages of (3.2) to generalizations such as the "functional efficiency" of Charnes and Cooper [5] will not be developed here.

First, by (3.2), the Pareto-optimal points (inputs!) among our n empirical points can be determined. The empirical Pareto-optimal function is then defined on the convex hull of these points by convex combinations of the "output" values. Note that the convex hull of the Pareto-optimal points might not include all of D_E since only the doubled line portion of the frontier is Pareto-optimal.

Since for efficient production we wish to maximize on outputs while minimizing on inputs, our relevant $g_k(x)$ include both outputs and inputs, e.g.

$$(3.3) \quad -g_k(x) \triangleq \begin{cases} y_k, & 1 \leq k \leq s \\ -x_i, & k = s+i, i=1, \dots, m \\ \text{for } (x,y) \in Q_E \end{cases}$$

For the optimization in (3.2) we clearly need only consider $(x,y) \in P_E$ rather than Q_E . Thus the constraint inequalities in (3.2) are for a test point (x^*, y^*) :

$$(3.4) \quad y \geq y^*, \quad x \leq x^*$$

and we have

Theorem 2: The envelopment constraints of Data Envelopment Analysis in production analysis are the Charnes-Cooper constraints for testing Pareto-optimality of an empirical production point.

In no way, as others, e.g. Färe [13] have mistakenly asserted, is Data Envelopment Analysis restricted to linear constant returns to scale functions or to truncated cone domains. Evidently via (3.2), Data Envelopment Analysis applies to much more general functions, function domains and other situations than the current empirical production function one.

To test an empirical "input-output" point (X_0, Y_0) for Pareto-optimality, the C^2 (Charnes and Cooper) test of (3.2) becomes

$$(3.5) \quad \begin{aligned} \min \quad & -e^T Y \lambda + e^T X \lambda \\ \text{subject to } & Y \lambda - s^+ = Y_0 \\ & -X \lambda - s^- = -X_0 \\ & e^T \lambda = 1 \\ & \lambda, s^+, s^- \geq 0 \end{aligned}$$

where $X \triangleq [X_1, \dots, X_n]$, $Y \triangleq [Y_1, \dots, Y_n]$.

Since $-e^T(Y\lambda - Y_0) + e^T(X\lambda - X_0)$ is an equivalent functional (it differs from the above one only by a constant), we can rewrite the problem for convenience

in later comparisons as:

$$\begin{aligned}
 & \min \quad -e^T s^+ - e^T s^- \\
 & \text{subject to } Y\lambda - s^+ = Y_0 \\
 (3.6) \quad & -X\lambda - s^- = -X_0 \\
 & e^T \lambda = 1 \\
 & \text{with } \lambda, s^+, s^- \geq 0
 \end{aligned}$$

This is the new DEA form for the production possibility set Q_E via P_E . As we shall see later, other variations of Q_E can be accommodated easily by simple modifications of or additions to the constraints on λ . Its informatics and software involve only minor modification from that of the Charnes, Cooper, Seiford and Stutz paper [11] as developed by I. Ali and J. Stutz for the Center for Cybernetic Studies of The University of Texas at Austin.

EFFICIENCY ANALYSIS

As mentioned, managerial and program comparison aspects of efficiency analysis were initiated by Charnes, Cooper and Rhodes in [6], [8], and [9], through a generalization of the single input, single output absolute efficiency determination of classical engineering and science to multi-input, multi-output relative efficiencies of a finite number of decision-making units "DMU's" (sometimes called "productive" units or "response" units). The multi-input, multi-output situations were reduced to the "virtual" single input single output ones through use of virtual multipliers and sums. Explicitly, the CCR ratio measure of efficiency of the DMU designated "o" is given by the non-linear, non-convex, non-Archimedean fractional program (see [7]).

$$\begin{aligned}
 & \text{Max} \quad \frac{n^T Y_o}{\xi^T X_o} \\
 & \text{subject to} \quad \frac{n^T Y_j}{\xi^T X_j} \leq 1 \quad , j = 1, \dots, n \\
 (4.1) \quad & - \frac{n^T}{\xi^T X_o} \leq -\epsilon e^T \\
 & - \frac{\xi^T}{\xi^T X_o} \leq -\epsilon e^T
 \end{aligned}$$

where the entries of the X_j and Y_j are assumed positive, ϵ is a non-Archimedean infinitesimal, e^T is a row vector of ones and, by abuse of notation, has s entries for n^T , m entries for ξ^T . (X_o, Y_o) is one of the n input-output pairs.

Employing the Charnes-Cooper transformation of fractional programming

$$(4.2) \quad \mu^T \triangleq \eta^T / \xi^T X_0, \quad v^T \triangleq \xi^T / \xi^T X_0, \quad v^T X_0 = 1$$

we obtain the dual non-Archimedean linear programs

$$(4.3) \quad \begin{array}{ll} \max & \mu^T Y_0 \\ \text{subject to} & v^T X_0 = 1 \\ & \mu^T Y - v^T X \leq 0 \\ & -\mu^T \leq \epsilon e^T \\ & -v^T \leq \epsilon e^T \end{array} \quad \begin{array}{ll} \min \theta & -\epsilon e^T s^+ - \epsilon e^T s^- \\ & Y\lambda - s^+ = Y_0 \\ & \theta X_0 - X\lambda - s^- = 0 \\ & \lambda, s^+, s^- \geq 0 \end{array}$$

where $X \triangleq [X_1, \dots, X_n]$, $Y \triangleq [Y_1, \dots, Y_n]$.

Although, clearly, no assumptions have been made concerning the type of functional relations for the input-output pairs (X_j, Y_j) , the dual program may be recognized as having the Data Envelopment Analysis constraints for an empirical production possibility set of Farrell, Shephard, etc. cone type $Q_E \cup B \cup C$, and, since

$$(4.4) \quad \theta = \epsilon[e^T Y\lambda - e^T X\lambda]$$

is an equivalent form for the functional, as being a Charnes-Cooper Pareto-optimality test for $(\theta X_0, Y_0)$ over the cone on the (X_j, Y_j) , $j=1, \dots, n$, with pre-emption on the intensity θ of input X_0 . As shown, for example, in [7], DMU_0 is efficient iff $\theta^* = 1$, $s^{*+} = 0$, $s^{*-} = 0$.

Re informatics, which are particularly important since all n efficiency evaluations must be made (i.e., n linear programs must be solved), the dual problem can be computed exactly (in the base field) as shown in [5],

e.g., with the code NONARC of Dr. I. Ali (Center for Cybernetic Studies, The University of Texas at Austin), or approximately by using a sufficiently small numerical value for ϵ . A typical efficient point is designated by $(\hat{x}, \hat{y})$ in Figure 1.

If a DMU is inefficient, the optimal $\lambda_j^* > 0$ in its DEA problem (=Charnes-Cooper test) designate efficient DMU's. Thus, a "proper" subset of the efficient DMU's determines the efficiency value of an inefficient DMU. The convex combinations of this subset are also efficient. Thereby to each inefficient DMU a "facet" of efficient DMU's is associated. The transformation

$$(4.5) \quad X_0 \rightarrow \theta^* X_0 - s^{*-}, \quad Y_0 \rightarrow Y_0 + s^{*+}$$

where the asterisk designates optimality, projects DMU₀, i.e., (X_0, Y_0) , onto its efficiency facet.

This projection was employed by Charnes, Cooper and Rhodes [9] to correct for differences in managerial ability in their analysis of programs Follow-Through and non-Follow-Through. It also shows quantitatively what improvements in inputs and outputs will (ceteris paribus) bring a DMU to efficient operation. Thus, although the relative efficiency measure of an inefficient DMU will involve the infinitesimal ϵ , non-infinitesimal changes for improvement are suggested.

Both Farrell and Shepard knew that ratio measures required adjustments to correctly exhibit inefficiency of the second DMU in examples like the following 2 input, 1 output, 2 DMU case:

DMU	x_1	x_2	y
1	1	2	1
2	1	4	1

Farrell added geometric points at infinity; Shephard simply excluded such cases without giving a method for their exclusion. The non-Archimedean extension in the CCR formulation is necessary to have an algebraically closed system of linear programming type. Linear programming theory holds for non-Archimedean as well as Archimedean entries in the vector and matrix problem data.

Our new Pareto-optimal DEA method like C^2S^2 [11] associates facets with non-optimal (=non-Pareto-efficient) DMU's. Clearly, by the C^2 -test, DMU_0 is Pareto-efficient (Pareto-optimal) iff $-e^T s^{*+} - e^T s^{*-} = 0$, i.e., iff the ℓ_1 -distance from (X_0, Y_0) to the farthest "northwesterly" (X_j, Y_j) point is zero. The CCR efficient DMU's are also among the new Pareto-optimal DMU's. Projection of a non-optimal DMU onto its Pareto-efficient facet is rendered by

$$(4.6) \quad X_0 \rightarrow X_0 - s^{*-}, \quad Y_0 \rightarrow Y_0 + s^{*+}$$

To achieve a convenient efficiency measure, we modify the functional by multiplying it by a $\delta > 0$ and consider

$$(4.7) \quad -\delta e^T s^{*+} - \delta e^T s^{*-},$$

where the asterisk denotes optimality, as the logarithm of the efficiency measure. When the data in X and Y are scaled to lie between 0 and 100, a $\delta = 1/10(m+s)$ will yield a logarithm between 0 and -10. This measure might then be called the "efficiency pH" by analogy with the pH of chemistry.

Our new measure relates to the units invariant multiplicative measure of Charnes, Cooper, Seiford and Stutz [12], which as shown there is necessary and sufficient that the DEA envelopments be piecewise Cobb-Douglas, by considering the entries in the X_j, Y_j to be logarithms of the entries in $\bar{X}_j, \bar{Y}_j$ which we employ in the multiplicative formulation.

INFORMATICS AND FUNCTION PROPERTIES

(A) Partial Derivatives:

The guidance provided by the CCR, BCC, C^2S^2 formulations does not include convenient access to the rates of change of the outputs with change in the inputs. The optimal dual variables in the DEA side linear programming problems give rates of change of the efficiency measure with changes in inputs or outputs. The non-Archimedean formulations further may give infinitesimal rates, which are not easily employed. And, for most of the efficient points one has non-differentiability because they are extreme points rather than (relative) interior points. Nevertheless, because of the informatics, e.g., computational tactics, we employ in testing via C^2 for Pareto-optimality, the following constructive method can be employed.

On reaching a non-Pareto optimal point, our software discovers all the optimal points in its facet, hence, implicitly, all the convex combinations which form the facet. Since the Pareto-optimal facet is a linear surface it is not only differentiable everywhere in its relative interior but all its partial derivatives are constant throughout the facet. Thus, we need only obtain these for any relative interior point of the facet to have them for the whole facet. Such a point is the average of the Pareto-optimal points of the facet.

Let

$$(5.1) \quad F(x_1, \dots, x_m, y_1, \dots, y_s) = 0$$

be the linear equation of the facet. Since we have sufficient differentiability in the neighborhood of the average point $(\bar{x}, \bar{y})$, we know

$$(5.2) \quad \left. \frac{\partial y_r}{\partial x_i} \right|_{\bar{x}, \bar{y}} = - \left(\frac{\partial F}{\partial x_i} \right) / \left(\frac{\partial F}{\partial y_r} \right)$$

where the right side partial derivatives are also evaluated at $(\bar{x}, \bar{y})$.

Suppose we run the C^2 -test with $(\bar{x}, \bar{y})$ as the point being tested. Then the optimal dual variables corresponding to input $\bar{x}_i$ and $\bar{y}_h$ are respectively

$$\left(\frac{\partial F}{\partial x_i} \right) \bigg|_{\bar{x}, \bar{y}} \quad \text{and} \quad \left(\frac{\partial F}{\partial y_h} \right) \bigg|_{\bar{x}, \bar{y}} . \quad \text{Thus, the rate of change of output } y_h \text{ with}$$

respect to input x_i is simply the negative of the ratio of the optimal dual x_i constraint variable to the optimal dual y_h constraint variable!

More specifically, all Pareto-optimal (X_j, Y_j) of the facet for the barycenter $_{\bar{x}, \bar{y}}^{(\bar{y})}$ satisfy

$$(5.4) \quad \mu^* T_y - v^* T_x - \varphi^* = 0$$

where (μ^*, v^*, φ^*) are the dual evaluators at an optimal basic solution, since they do not depend on the C^2 -test right hand sides. Thereby our

$$(5.5) \quad F(x, y) = \mu^* T_y - v^* T_x - \varphi^* = 0$$

Clearly, $\mu_h^* = \partial F / \partial y_h$, $-v_i^* = \partial F / \partial x_i$ as already stated.

(B) Isotonicity and Economies of Scale:

Theorem 1 shows that every component of the empirical frontier production function is a concave function.

Suppose x^1 and x^2 are the inputs of two Pareto-optimal DMU's in the same facet and $x^1 \geq x^2$. Since $x^1 = \chi \lambda^*(x^1)$ and $x^2 = \chi \lambda^*(x^2)$ we must have $e^T \chi \lambda^*(x^1) \geq e^T \chi \lambda^*(x^2)$. But for Pareto-optimality, $e^T \gamma \lambda^*(x^i) = e^T \chi \lambda^*(x^i)$, $i=1,2$ so that $e^T \gamma \lambda^*(x^1) \geq e^T \gamma \lambda^*(x^2)$. Then, letting $f^P(x)$ denote the empirical Pareto-optimal (vector) function we have

$$(5.3) \quad e^T f^P(x^1) \geq e^T f^P(x^2)$$

Further, if $x^\mu \triangleq \mu x^1 + (1-\mu)x^2$, $0 \leq \mu \leq 1$, $f^P(x^\mu) \leq \mu f^P(x^1) + (1-\mu)f^P(x^2)$ by construction of the empirical frontier function and we have

$$e^T f^P(x^1) \geq \mu e^T f^P(x^\mu) \geq e^T f^P(x^2).$$

For the single output case of Farrell, etc., then

Theorem 3: If there is only a single output, the empirical Pareto-optimal production function is isotonic in every facet (regardless of what underlying production function we have sampled from).

Proof: A function $f(x)$ is "isotonic" iff $x^a \geq x^b$ implies $f(x^a) \geq f(x^b)$.

Also $e^T f^P(x) = f^P(x)$ with a single output.

Possibly because of ignorance of standard mathematical terminology, the isotonic property has been called "strong disposability" in the economics literature. The name "weak disposability" has also been used for the weaker property $f(\rho x) \geq f(x)$ whenever $\rho \geq 1$. A better name might be "ray isotonic."

Our arguments preceding Theorem 3 establish a "sum isotonic" property on facets for the empirical Pareto-optimal function with multiple output components (regardless of the underlying production function set we have sampled from), namely,

Theorem 4: $e^T fP(x^a) \geq e^T fP(x^b)$ whenever $e^T x^a \geq e^T x^b$ with x^a, x^b in the same facet.

Classically in economics, production functions studied have usually been assumed to be homogeneous and defined on the non-negative orthant. Thereby, whether or not a function for which $f(\rho x) = \rho^\alpha f(x)$, with $\rho \geq 0$, had economies of scale would be decided by the value of the exponent α . More generally, increasing or decreasing "return to scale" would be present respectively, at $\bar{x}$ if $f(\rho \bar{x}) > \rho f(\bar{x})$ or $f(\rho \bar{x}) < \rho f(\bar{x})$ for $\rho > 1$ at points $\bar{x}$ in a small neighborhood of $\bar{x}$. The BCC paper [3] gives a criterion for deciding this (with production possibility set $Q_E \cup B \cup C$ or $Q_E \cup B$) but does not give us the rates of change.

Because of our preceding theorems, however, we know that empirical Pareto-optimal functions are sum-isotonic on facets and concave in each component function regardless of the nature of the underlying production possibility set. Thereby, we automatically anticipate lower and lower returns to scale in going from facet to facet with increasing $e^T x$. And our partial derivatives can give us explicitly the rates of change in each observed facet.

Practically, our choices of inputs are generally made with the expectation that the underlying Pareto-optimal function is isotonic, i.e., we choose the form of the inputs so that an increase in an input should not decrease the outputs. But even here we need still more to determine the non-concave portions of an isotonic function. For example, in Figure 2 an isotonic function is plotted together with the resulting concave cap

(large dashed lines) obtained as the empirical function:

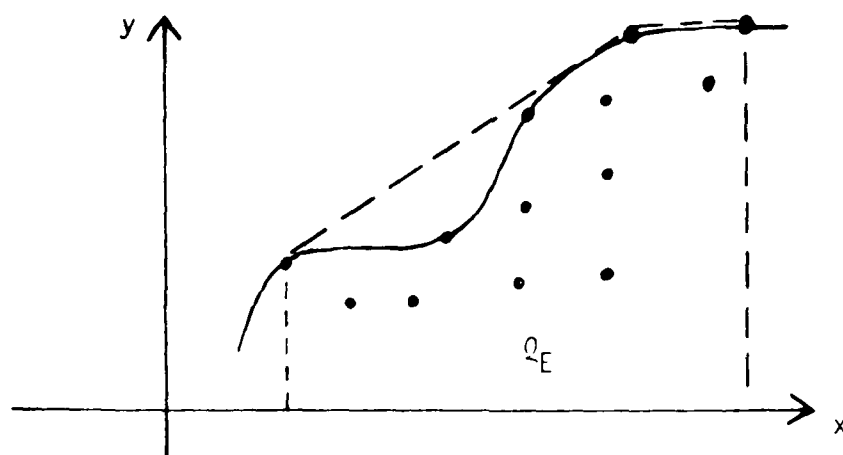


Figure 2

As suggested in our original (1981) paper, non-concavity can be explored by applying (output) component by component strictly concave transformations g_t to obtain $g_t(y_t)$ instead of y_t so that $g_t(y_t(x))$ would be concave and our plot would look like

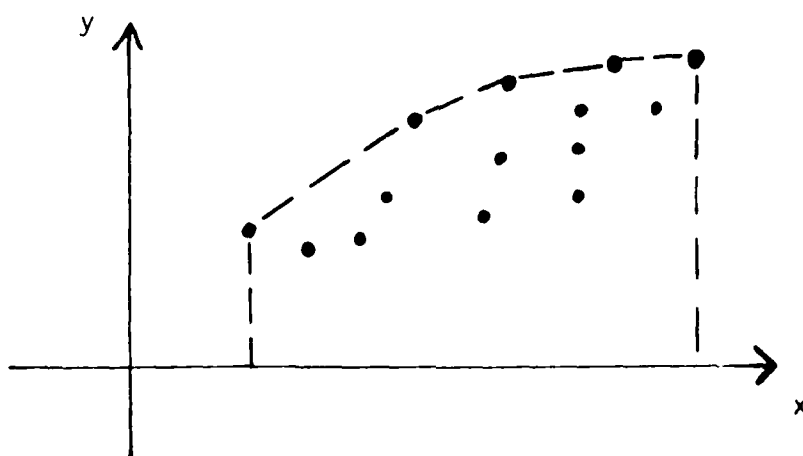


Figure 3

(C) Discretionary and Non-Discretionary Inputs:

In a number of practical applications, certain relevant inputs, e.g., unemployment rate, population, median income, are not subject to "discretionary" change by the decision-makers of decision-making units. These are called "non-discretionary" inputs. They are important in influencing the outputs and in furnishing the reference background in terms of which units' efficiency is rated. Not infrequently the facet associated with an inefficient unit has the same values for the non-discretionary inputs, in which case there is no problem with the rating assigned. If not, however, to obtain more meaningful ratings we can add constraints on λ to those in (3.5) which require the non-discretionary inputs to be the same as that of the unit being evaluated. Thereby, a more meaningful rating will be attained.

CONCLUSIONS

We have shown how direct application of the Charnes-Cooper test for Pareto optimality leads to a simpler and more robust method, efficiency PH , encompassing all previous ones for ascertaining "efficiency." Further, Pareto-optimal characterizations and constructions of empirical production functions restrict us methodologically to exploration of such functions by means of concave sum-isotonic caps. Economies of scale from these thereby expectedly decrease with increase in the magnitude of the input vectors. Use of transformations of outputs, as we suggest, can uncover non-concave regions of the underlying production function where substantial economies of scale may prevail. Our new informatics device and theory of the use of the facet average (or barycenter) also constructively furnishes quantitative estimates of the rates of change of outputs with respect to inputs which have not been available previously. These new devices, as with other usages of empirical functions, suggest important new areas for development of statistical theory to distinguish between true properties and sampling "accidents." The vital importance of further development of the informatics of solution of systems of adaptively developed linear programming problems for Pareto-optimal constructions should also be clear.

Informatically, we are doing this by applying transformations of form $g_h(y_h) = \bar{y}_h + (y_h - \bar{y}_h)^{\frac{1}{\beta}}$ with $\beta \approx 20$ to obtain possible new facets in the $g_h(y_h)$.

Problems do arise, of course, on whether one gets spurious empirical frontier portions in this manner for empirical points which should "really" be inefficient. Evidently such non-concave portions are portions of increasing returns to scale if they are truly on the frontier.

REFERENCES

1. Afriat, S., "Efficiency Estimation of Production Functions," International Economic Review, 13, 3, 568-598 (1972).
2. Aigner, D.J., C.A.K. Lovell and P.J. Schmidt, "Formulation and Estimation of Stochastic Frontier Production Function Models," Journal of Econometrics, 6, 1, 21-37 (1977).
3. Banker, R., A. Charnes and W.W. Cooper, "Estimation of Technical and Scale Efficiencies in Data Envelopment Analysis," Management Science (forthcoming).
4. Ben-Israel, A., A. Ben-Tal and A. Charnes, "Necessary and Sufficient Conditions for a Pareto Optimum in Convex Programming," Econometrica, 45, 4, 811-820 (1977).
5. Charnes, A. and W.W. Cooper, Management Models and Industrial Applications of Linear Programming, J. Wiley and Sons (1961).
6. Charnes, A. and W.W. Cooper, "Auditing and Accounting for Program Efficiency in Not-For-Profit Entities," Accounting, Organizations and Society, 5, 1, 87-107 (1980).
7. Charnes, A., W.W. Cooper, A. Lewin, R. Morey and J. Rousseau, "An Approach to Positivity and Stability Analysis in DEA," Management Science (forthcoming).
8. Charnes, A., W.W. Cooper and E. Rhodes, "Measuring the Efficiency of Decision-Making Units," European Journal of Operations Research, 2, 429-444 (1978).
9. Charnes, A., W.W. Cooper and E. Rhodes, "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through," Management Science, 27, 668-697 (1981).
10. Charnes, A., W.W. Cooper and A. Schinnar, "Transforms and Approximations in Cost and Production Function Relations," Omega, 10, 2, 207-211 (1982).
11. Charnes, A., W.W. Cooper, L. Seiford and J. Stutz, "A Multiplicative Model for Efficiency Analysis," Socio-Economic Planning Sciences, 16, 5, 223-224 (1982).
12. Charnes, A., W.W. Cooper, L. Seiford and J. Stutz, "A Units Invariant Multiplicative Measure," Operations Research Letters, July 1983.
13. Färe, R. and C.A.K. Lovell, "Measuring the Technical Efficiency of Production," Journal of Economic Theory, 19, 1, 150-162 (1978).

14. Farrell, M.J. "The Measurement of Productive Efficiency," J. Royal Statistical Society, Series A, General, 120, 3, 253-281 (1957).
15. Fenchel, W., Convex Sets, Cones and Functions, Princeton University Press, Princeton, N.J. (1953).
16. Försund, F. and L. Hjalmarsson, "Generalized Farrell Measures of Efficiency: An Application to Milk Processing in Swedish Dairy Plants," Economic Journal, 89, 294-315 (1979).
17. Rockefellar, T., Convex Analysis, Princeton University Press, Princeton, N.J. (1970).
18. Shephard, R.W., Cost and Production Functions, Princeton University Press, Princeton, N.J. (1953).
19. Shephard, R.W., Theory of Cost and Production Functions, Princeton University Press, Princeton, N.J. (1970).

III. INVARIANT MULTIPLICATIVE EFFICIENCY AND PIECEWISE COBB-DOUGLAS ENVELOPMENTS

Introduction

In [1], Charnes, Cooper, Seiford, and Stutz (C²S²) develop a multiplicative (or log) measure of the relative efficiency of multiple input, multiple output productive (or "decisionmaking") units (DMU's). In contrast to the CCR measure [2, 3], the multiplicative measure obtained in [1] is not invariant under change of units in the inputs or outputs. We show here how by a simple change preserving the multiplicative format that a units invariant multiplicative measure can be obtained. Interestingly, the Data Envelopment Analysis (DEA) associated with this new modification necessarily yields optimal envelopments by Cobb-Douglas functions, i.e., the efficiency surface is piecewise Cobb-Douglas rather than merely log-linear! This uncovers a new role for Cobb-Douglas functions¹--they are necessary for the units invariant property of a multiplicative measure.

Units Invariant Multiplicative Efficiencies

The C²S² multiplicative model reduces the input-output quantities to single virtual output to input ratios. If we now introduce an additional virtual output multiplier and virtual input multiplier, we obtain the following form for our problem to measure the efficiency of DMU_j relative to all the n DMU's:

$$\begin{aligned}
 & \max (e^{\hat{\eta}} \prod_{r=0}^s y_{r0}^{\mu_r}) / (e^{\hat{\xi}} \prod_{i=1}^m x_{i0}^{\nu_i}) \\
 (1) \quad & \text{s.t.} \quad (e^{\hat{\eta}} \prod_{r=0}^s y_{rj}^{\mu_r}) / (e^{\hat{\xi}} \prod_{i=1}^m x_{ij}^{\nu_i}) \leq 1, \quad j = 1, \dots, n \\
 & -\hat{\eta} \leq 0, \quad -\hat{\xi} \leq 0, \quad -\mu_r \leq -\delta, \quad -\nu_i \leq -\delta, \quad \forall r, i.
 \end{aligned}$$

¹Other properties relating Cobb-Douglas forms to more general classes of functions are examined in [5].

where $\delta > 0$ and DMU_0 is one of the n DMU's in the constraints.

Suppose the units in the outputs and the inputs are changed so that Y_{rj} becomes $a_r Y_{rj}$ and X_{ij} becomes $b_i X_{ij}$ where $a_r, b_i > 0, \forall r, i$ (Note a_r or $b_i = 1$ corresponds to no change in those units). Problem (1) becomes

$$\begin{aligned}
 (2) \quad & \max \quad e^{\hat{\eta}(\prod_r a_r^{\mu_r}) \prod_r Y_{r0}^{\mu_r}} / e^{\hat{\xi}(\prod_i b_i^{v_i}) \prod_i X_{i0}^{v_i}} \\
 & \text{s.t.} \quad e^{\hat{\eta}(\prod_r a_r^{\mu_r}) \prod_r Y_{rj}^{\mu_r}} / e^{\hat{\xi}(\prod_i b_i^{v_i}) \prod_i X_{ij}^{v_i}} \leq 1, j = 1, \dots, n \\
 & \quad \hat{\eta} \leq 0, \quad \hat{\xi} \leq 0, \quad -\mu_r \leq -\delta, \quad -v_i \leq -\delta, \quad \forall r, i.
 \end{aligned}$$

If (1) has optimal value $E(1)$ with $\eta^*, \xi^*, \mu^*, v^*$ an optimal solution, then $\exp(\hat{\eta}) = K \exp(\hat{\eta}^*) / \prod_r a_r^{\mu_r^*}$, $\exp(\hat{\xi}) = K \exp(\hat{\xi}^*) / \prod_i b_i^{v_i^*}$, μ^*, v^* (where $K > 0$ assures $\hat{\eta}, \hat{\mu} > 0$) is "feasible" for (2) with value $E(1)$. Hence for (2) the optimal value $E(2) \geq E(1)$. Similarly from an optimal solution $\bar{\eta}, \bar{\xi}, \bar{\mu}, \bar{v}$ to (2) we construct a feasible solution to (1) with value $E(2)$. Thereby $E(1) \leq E(2) \leq E(1)$ i.e., the efficiency value is invariant under change of units.

The Cobb-Douglas Property

Taking logarithms in (1) and going to vector matrix notation as in [1], we obtain the dual linear programming problems:

$$\begin{aligned}
 (3) \quad & \begin{array}{ll} \text{I} & \text{II} \\ \max \quad \hat{\eta} & - \hat{\xi} + \mu^T \hat{Y}_0 - v^T \hat{X}_0 \\ \text{s.t.} \quad \hat{\eta} e^T - \hat{\xi} e^T + \mu^T \hat{Y} - v^T \hat{X} & \leq 0 \\ & -\hat{\eta} \leq 0 \\ & -\hat{\xi} \leq 0 \\ & -\mu^T \leq -\delta e^T \\ & -v^T \leq -\delta e^T \end{array} & \begin{array}{ll} \min & -\delta e^T s^+ - \delta e^T s^- \\ \text{s.t.} & e^T \lambda - \theta^+ = 1 \\ & -e^T \lambda - \theta^- = -1 \\ & \hat{Y} \lambda - s^+ = \hat{Y}_0 \\ & -\hat{X} \lambda - s^- = -\hat{X}_0 \\ & \lambda, \theta^+, \theta^-, s^+, s^- \geq 0 \end{array} \end{array}$$

Here II represents the DEA side of the efficiency problem. Adding the first two equations in II, we obtain $-\theta^+ - \theta^- = 0$. Since $\theta^+, \theta^- \geq 0$ we must have $\theta^+ = \theta^- = 0$. Thus II reduces to

$$(4) \quad \begin{aligned} \min \quad & -\delta e^T s^+ - \delta e^T s^- \\ \text{s.t.} \quad & \hat{Y}\lambda - s^+ = \hat{Y}_0 \\ & -\hat{X}\lambda - s^- = -\hat{X}_0 \\ & e^T \lambda = 1 \\ & \lambda, s^+, s^- \geq 0 \end{aligned}$$

Thereby we have $\hat{Y}_0$ and $\hat{X}_0$ enveloped by convex combinations of the $\hat{Y}_j, \hat{X}_j$. With optimal solutions $\lambda^*, s^{*+}, s^{*-}$, we can write

$$(5) \quad \begin{aligned} Y_0 &= \prod_{j=1}^n Y_j^{\lambda_j^*} e^{-s_j^{*+}} \\ X_0 &= \prod_{j=1}^n X_j^{\lambda_j^*} e^{s_j^{*-}} \end{aligned} \quad \text{where } \sum_j \lambda_j^* = 1$$

and by $Y_j^{\lambda_j^*}$, resp. $X_j^{\lambda_j^*}$, we mean $(Y_{1j}^{\lambda_j^*}, \dots, Y_{sj}^{\lambda_j^*})^T$ resp. $(X_{1j}^{\lambda_j^*}, \dots, X_{mj}^{\lambda_j^*})^T$.

Thus our optimal envelopments are by Cobb-Douglas functions with $\lambda_j^* > 0$ implying that DMU_j is efficient, i.e., DMU_0 is associated with the efficiency surface "facet" spanned by those DMU_j 's for which $\lambda_j^* > 0$.

We note further that the simplified dual programs corresponding to (3) are now

$$(6) \quad \begin{array}{ll} \begin{array}{l} \text{(I')} \\ \max \quad \mu^T \hat{Y}_0 - v^T \hat{X}_0 + \omega \\ \text{s.t.} \quad \mu^T \hat{Y} - v^T \hat{X} + \omega e^T \leq 0 \\ \quad \quad -\mu^T \leq -\delta e^T \\ \quad \quad -v^T \leq -\delta e^T \end{array} & \begin{array}{l} \text{(II')} \\ \min \quad -\delta e^T s^+ - \delta e^T s^- \\ \text{s.t.} \quad \hat{Y}\lambda - s^+ = \hat{Y}_0 \\ \quad \quad -\hat{X}\lambda - s^- = -\hat{X}_0 \\ \quad \quad e^T \lambda = 1 \\ \quad \quad \lambda, s^+, s^- \geq 0 \end{array} \end{array}$$

These results present us with a new method for estimating piecewise Cobb-Douglas production functions directly from empirical data. The form of (II') in contrast to that of [4] is also sufficiently simple that one can anticipate that the mathematical statistics of this type of Cobb-Douglas estimation may well be developed in the near future (see also the Appendix in [3]).

REFERENCES

- [1] Charnes, A., Cooper, W. W., Seiford, L., and Stutz, J., "A Multiplicative Model for Efficiency Analysis," Socio-Economic Planning Sciences (forthcoming 1983); also CCS 416, Center for Cybernetic Studies, The University of Texas, Austin, Texas, November 1981.
- [2] Charnes, A., Cooper, W. W., and Rhodes, E., "Measuring the Efficiency of Decisionsmaking Units," European Journal of Operational Research, v. 2, pp. 429-444, 1978.
- [3] Charnes, A. and Cooper, W. W., "Management Science Relations for Evaluation and Management Accountability," Journal of Enterprise Management, v. 2, pp. 143-162, 1980.
- [4] Banker, R. D., Charnes, A., Cooper, W. W., and Schinnar, A. P., "A Bi-Extremal Principle for Frontier Estimation and Efficiency Evaluations," Management Science, v. 17, No. 12, pp. 1370-1382, December 1981.
- [5] Charnes, A., Cooper, W. W., and Schinnar, A. P., "A Theorem on Homogeneous Functions and Extended Cobb-Douglas Forms," Proc. Natl. Acad. Sci., U.S.A., 73, No. 10, pp. 3747-3748, October 1976.

IV. A COMPARITIVE STUDY OF DATA ENVELOPMENT ANALYSIS AND OTHER APPROACHES TO EFFICIENCY EVALUATION AND ESTIMATION

1. Introduction

Data Envelopment Analysis (DEA) is a new efficiency measurement methodology developed by A. Charnes, W. W. Cooper, and E. Rhodes as set forth in [12] [13] and [14].^{1/} It is designed to measure the relative efficiency of Decision Making Units (DMUs) which use multiple inputs to produce multiple outputs even when the underlying production function is not known and where, additionally, these functions may also be multiple in character. This contrasts with the situation for statistical techniques and theory, e. g., as employed in economics, where either the underlying production function must be known, or at least its parametric form must be assumed before it can be used to evaluate efficiencies and where, usually, a single functional form is also assumed. See, e. g., Feldstein [18]. See also [32] and [33]. The latter, regression approaches, are thus limited, especially in the case of public sector institutions such as hospitals, etc., where programs and activities are even less readily identified for such assumptions than is the case in industrial production.

DEA has now been applied to several types of organizations including education [5] [6], health care [4] [29], Navy recruiting [22], and criminal court systems [21]. Nevertheless something more is required and, in particular, the validity and reliability of DEA in locating inefficient DMUs, identifying the inputs (and/or outputs) where the inefficiencies occur and estimating their amounts or magnitudes all need to be evaluated. One way to approach this task is via a situation in which the identity of the truly inefficient units is known along with the sources and amounts of this inefficiency. This paper therefore attempts to evaluate DEA through use

^{1/} See also [25].

of an artificial data base where the efficient and inefficient DMUs are all known in numerical detail. DEA's performance is then compared with other commonly employed techniques such as ratio and regression analyses.

Regression and ratio analyses were selected for these evaluations because they are widely used in fields like health services, which is the field we shall use to guide our data base construction. In this paper we restrict our examination only to some of the fairly simple forms of ratio and/or regression approaches that are in wide use.^{1/} More sophisticated regression techniques such as the translog function and other so-called "flexible functional form" approaches are considered elsewhere. See Sherman [29].

The following section describes how the data base was constructed and section 3 discusses the data base that was developed. Section 4 describes the version of DEA that will be used while sections 5, 6 and 7 discuss the results of applying DEA, ratio and regression analyses to this data base. The resulting comparisons are summarized in section 7 with respect to the ability of these techniques to identify and distinguish between efficient and inefficient DMUs. Section 8 then extends the uses of DEA to locating and estimating the amounts of inefficiencies in particular DMUs in ways that are not generally available when the ratio or regression approaches are used. A concluding section then discusses some of the shortcomings found in these other approaches and indicates where they differ from DEA and how some of their shortcomings might be repaired.

^{1/} Similarly, only one version of DEA is used and no attempt is made to distinguish between various types of efficiencies such as scale vs. technical efficiencies and other sources of inefficiency such as are examined in [3]. Finally, we did not use statistical techniques to develop our data base, as was done in [2], and hence can make only limited use of statistical significance tests and like devices for generalizing our results. Our purpose is rather to supply insight of potential value on the use of the techniques we study rather than to secure generalizations for the different data situations that might be encountered in actual practice.

2. Model Structure and Data Generation

The artificial data set was constructed by defining a hypothetically "known" technology which applies to all Decision Making Units (DMUs) and defines efficient input-output relationships for each of them.^{1/} Inefficiencies which were explicitly introduced for certain DMUs take the form of excess inputs used for the output levels attained. Hence, a DMU that achieves its output level by using only the amount of inputs required by this hypothetical technology is efficient while a DMU that uses more than the required amount of any input is inefficient. To make the inputs and outputs easier to recognize, they are referred to and labelled in the context of a hospital study as one area of potential interest. See Sherman [29]. We assume that these hospitals are all public (not-for-profit) institutions so that the usual profit calculus and/or price-weighted reductions to a scalar measure of efficiency evaluation are not wholly appropriate.

^{1/} Knowledge gained from the study of Massachusetts hospitals reported in [29] was used in the choice of inputs and outputs and in the construction of the data set.

The set of artificial hospital data generated for our simulation consisted of three outputs produced with three inputs during a one year period of time^{1/} as follows:

<u>Outputs</u>	<u>Inputs</u>
y ₁ : Regular patient* care/year (patients treated in one year with average level of inputs for treatment)	x ₁ : Staff utilized in terms of full-time equivalents, i.e., (FTE s)/year
y ₂ : Severe patient* care/year (patients treated in one year with severe illness requiring higher input levels than regular patients for more complex treatment).	x ₂ : Number of hospital bed days available/year
y ₃ : Teaching of residents and interns/year (number of individuals receiving one year of training)	x ₃ : Supplies in terms of dollar cost/year

*measured in terms of number of patients treated

The data set to be generated was for 15 hypothetical hospitals which we label as H1, H2, ..., H15, to represent the pertinent DMUs.^{2/} They are all assumed to achieve their outputs via a common production process, which they may use efficiently or inefficiently. The resulting observed values are then constructed in a manner that we shall shortly describe.

In this study we shall focus on input inefficiencies, by which we mean that one or more of the above inputs may be used in excess to obtain a particular hospital's output values. Although we could also similarly study output deficiencies (in the form of output shortfalls from given inputs)^{3/} we shall not lengthen the paper to undertake that study here. In any case the known values of the per unit inputs for efficient production are given in Exhibit 1 inserted at the end of this paper.

^{1/}I. e., we are considering all data as annual rates.

^{2/}Subdivisions may also be used such as, e. g., the surgical units within each hospital that were studied in [29].

^{3/}An output shortfall approach from given inputs is used in [2].

The usual regression approach to efficiency and related types of economic analyses in multiple output situations uses a single aggregate function of a linear or logarithmic variety in which total cost is regressed against the observed output values. See, e. g., [18]. This approach carries with it a variety of assumptions^{1/} which we shall try to favor in our construction by using the same prices and a common technology for all DMUs. We shall not assume that all DMUs operate on their efficiency frontiers, however, but we shall otherwise proceed in accordance with the usual methods of estimation, testing and analyses that have been commonly employed in regression studies of health services and related fields.

To make the sense of this discussion more precise, we present our expressions for generating the inputs required for efficient operations by any hospital in the following form:

$$x_{ij} = \sum_{r=1}^3 a_{irj} y_{rj} \quad (1)$$

where

x_{ij} = amount of input i used per year by hospital j

y_{rj} = amount of output r produced per year by hospital j

a_{irj} = amount of input i used per unit of output r by
hospital j during the year.

^{1/} See, e. g., Sato [27].

^{2/} A use of DEA to distinguish coefficients for input-output analyses derived from data for efficient and inefficient sets of operations may be found in Schinnar [28].

These a_{irj} values, which are fixed constants, represent an efficient set of coefficients which may be used to generate the inputs required for any observed (or planned) level of outputs. In some cases we will assign values $\hat{a}_{irj} > a_{irj}$ for some i , r and j to represent managerial (= hospital) inefficiencies which yield values

$$\hat{x}_{1j} = \sum_{r=1}^3 \hat{a}_{irj} y_{rj}, \quad (2)$$

with $\hat{x}_{1j} > x_{1j}$ when inefficiencies are present.

The efficient a_{ir} values are given, free of any of the $j = 1, \dots, 15$ hospital identification subscripts, in Exhibit 1. These values are the same for all hospitals so that $a_{11} = .004$ FTE/patient represents the efficient labor requirement in Full Time Equivalent units per regular patient. Similarly $a_{12} = .005$ FTE/patient represents the efficient requirement for a severe patient and $a_{13} = .03$ FTE/training unit represents the efficient requirement to train one new resident/intern during a year.

Analogous remarks apply to the values $a_{21} = 7$ bed days/patient, and $a_{22} = 9$ bed days/patient for regular and severe patients, respectively, shown in the Bed Days column of Exhibit 1. The blank shown in the row for Training Units in this column means that $a_{23} = 0$ applies. That is, no Bed Days enter into the training outputs.

Finally, $a_{31} = \$20$ /patient and $a_{32} = \$30$ /patient represent the efficient level of supplies required per regular and severe patients, respectively, while $a_{33} = \$500$ /training unit is the coefficient for efficient training operations in output $r = 3$. Putting this $i = 3$ input in dollar units avoids the detail that would otherwise be needed to identify the different types of supplies that would be required for teaching and for different types of patient treatments.

DEA does not require reductions to cost equivalents. The various outputs and inputs may be specified in different units of measure and, indeed, it can be shown that the resulting DEA efficiency value is independent of the units of measure used in any output or input.^{1/} On the other hand reductions like these are required for the ratio and regression measures we shall also study. Therefore we next show how the efficient costs are derived to obtain this part of our data set. This is done via expressions of the form,

$$c_r = \sum_{i=1}^3 k_i a_{ir} \quad r = 1, 2, 3, \quad (3)$$

where we have omitted the index j for hospital identification because only efficient costs are being considered. Here k_i represents the cost of the i^{th} input requirement for the r^{th} output under efficient operations where

$$\begin{aligned} k_1 &= \$10,000/\text{FTE} \\ k_2 &= \$10/\text{bed day} \\ k_3 &= \$1/\text{supply unit}. \end{aligned} \quad (4)$$

These data are then combined with the preceding a_{ij} values to obtain

$$\begin{aligned} c_1 &= k_1 a_{11} + k_2 a_{21} + k_3 a_{31} = \$130/\text{regular patient} \\ c_2 &= k_1 a_{12} + k_2 a_{22} + k_3 a_{32} = \$170/\text{severe patient} \\ c_3 &= k_1 a_{13} + k_2 a_{23} + k_3 a_{33} = \$500/\text{training unit}. \end{aligned} \quad (5)$$

These are the formulas used at the bottom of Exhibit 1 to produce the efficient cost of outputs shown in the last column in the body of the table.

^{1/} Provided, of course, that these same units of measure are used for the specified output (or input) in the data for every DMU. See Charnes, Cooper and Rhodes [10]. See also Rhodes [25] and Charnes and Cooper [7].

3. Data Base Development

We now turn to Exhibit 2 which reflects the composition of inefficient and efficient hospitals included in our data base. The hypothesized "actual" (or observed) inputs per unit output used by each hospital, whether efficient or not, are listed in Exhibit 2, columns 9-16 with inefficient input levels per unit of output denoted by . Column 17 reflects the actual vacancy rate (% of unused bed days available during the year) where, as noted in Exhibit 1, an efficient hospital is expected to have a 5% vacancy rate.

We develop the actual inputs used for each hospital in the manner we have already described by first selecting an arbitrary set of output values for each of the hospitals listed in the left-hand stub.^{1/} Teaching units per year are reflected in column 6, regular patients treated during the year are in column 7, and severe patients treated during the year are in column 8.

Other ways of summarizing patient care outputs for later use are included in columns 4 and 5. Column 4 reflects total patients as the sum of column 7 and column 8. Column 5 reflects the percentage (%) of severe patients treated which is based on $(\text{column 8}) \div (\text{column 4}) \times (100)$. We develop this percentage output measure because it reflects output data in a form which is often used to evaluate efficiency in many real data sets.^{2/}

The inputs used by each hospital to produce the outputs in columns 6, 7, and 8 are reflected in columns 1, 2, and 3. Column 1 contains the full time equivalents (FTE s) of labor years used. Column 2 has the bed days/year which were available and column 3 gives the supply dollars used during the year.

^{1/} Although these values could have been selected by statistical principles--e.g., of an experimental design variety--there seemed to be little point in doing so because our objective was to secure insight rather than the kinds of generalizability that require statistical tests of significance. See [2], however, for a study of the latter type.

^{2/} See the discussion in Sherman [29].

The values in columns 1, 2, 3 reflect mixtures of efficient and inefficient utilization of resources because of the way they were derived. We can clarify this by means of Exhibit 3 which illustrates how the data for H1, an efficient DMU, and H15, an inefficient DMU, were constructed. H1 is efficient and therefore used the same inputs per unit outputs as the structural model in Exhibit 1. During the year, H1 provided care for 3000 regular patients, 2000 severe patients, and 50 training units of service. It therefore utilized $(.004)(3000) + (.005)(2000) + (.03)(50) = 23.5$ FTEs in that year. H15 produced the same outputs as H1 but was inefficient in its use of certain inputs. It used .005 FTEs /regular patient, while it adhered to the structural model FTE usage rates for severe patients (.005 FTEs /patient) and training (.03 FTEs /training unit). H15 therefore used $(.005)(3000) + (.005)(2000) + (.03)(05) = 26.5$ FTEs /year to produce the same outputs. Similarly, H15 is inefficient in the number of bed days used and supply dollars used per regular patient but is efficient in the amount of bed days and supply dollars consumed for severe patients and for supply dollars used for teaching outputs. Bed days and FTEs and supply dollar inputs are also calculated in Exhibit 3 to further illustrate the way the data base was constructed.

The number of FTEs , bed-days, and supply dollars inputs were calculated as illustrated in Exhibit 3 for each hospital based on the arbitrarily assigned output mix of regular patients, severe patients and training units and actual efficient or inefficient input per unit output rate reflected in Exhibit 2.

Certain relationships posited in the structural model are generally not known, like the actual amount of staff time and supplies that are required to support each intern or resident at a hospital. We nevertheless explicitly introduce these relationships to determine if the efficiency measurement techniques we will apply can uncover them. Before proceeding, however, it should perhaps be noted that when the underlying structural model is known, the determination of which DMUs are inefficient can be directly determined and techniques such as we will be considering would be unnecessary for purposes of efficiency evaluation.

4. The DEA Model:

The Charnes Cooper Rhodes (CCR) model for data envelopment analysis which we will use assumes the following form:

Objective:

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

Constraints:

(6)

$$\begin{array}{l} \text{Less than} \\ \text{Unity} \\ \text{Constraints} \end{array} : 1 \geq \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m w_i x_{ij}} ; j = 1, \dots, 15$$

$$\begin{array}{l} \text{Positivity} \\ \text{Constraints} \end{array} : \begin{array}{l} 0 < u_r ; r = 1, \dots, s \\ 0 < w_i ; i = 1, \dots, m \end{array}$$

Data:

Outputs: y_{rj} = observed amount of r^{th} output for j^{th} hospital

Inputs: x_{ij} = observed amount of i^{th} input for j^{th} hospital.

1/ Other models which might have been used can be found in [3] and [15].
See also [16].

This model is therefore in fractional programming form with fractional constraints. As noted in Charnes, Cooper and Rhodes [13] it may be replaced by an ordinary linear programming model that also has non-Archimedean conditions imposed on the variables for what are here referred to as positivity constraints.^{1/}

We shall not enter into this kind of development but shall instead try to explicate what is happening in our DEA analysis by means of the above model. First we observe that the efficiency ratings are all restricted to an upper limit of unity. One of the $j = 1, \dots, 15$ hospitals, when singled out for efficiency evaluation, is represented in the objective as well as the constraints. By virtue of the latter condition we must have $\max h_o = h_o^* \leq 1$. Furthermore all observations y_{ij} and x_{ij} are positive so that, together with the positivity imposed on the variables, we will also have $0 \leq h_o^* \leq 1$ with $h_o^* = 1$ when and only when DMU_o , the DMU being evaluated, is efficient.

Qualifications need to be entered to allow for the presence of slack in the corresponding linear programming model.^{2/} We will not treat this topic in rigorous detail in the present paper but will instead supply an illustration with accompanying discussion that will provide insight into what is involved. Here we need only say that when slack is present in some input then, with efficiency, that input may be reduced to a new input level by

^{1/} See Charnes, Cooper, Lewin, Morey and Rousseau [11] for a precise development.

^{2/} Any slack which occurs in (6) is simply the complement of an efficiency rating but the development in [11] provides a way of identifying the presence of non-Archimedean values in (6) with slack in the corresponding linear programming model.

removing the slack without affecting any output or any other input. Hence the input which involved this slack was excessive and the operation could not have been efficient.

Bearing this in mind we next initiate our DEA analysis by reference to the data of Exhibit 2 after which we shall attempt to compare the resulting efficiency ratings with cost ratio and regression approaches applied to this same data base.

5. Applications to Artificial Data Base.

Applying (6) to Exhibit 2 with each of H1,...,H15 inserted in the objective produces the h_0^* values reported in Table 1. Every one of the efficient DMU's has received a rating of $h_0^* = 1$ but two inefficient DMU's--H10 and H13--are also accorded a value of $h_0^* = 1$ even though they are inefficient. The six DMU's that are rated as inefficient, with $h_0^* < 1$, are accorded these values by comparison with certain efficient units that comprise an efficiency reference set for the inefficient DMU (see Table 1). For example, H8 was found to be inefficient by direct comparison with H4; and H15 is being compared directly with H4, H6, and H7. This reference set, we need only note here, is supplied as part of the optimum basis in the linear programming computations. Hence the model and computing routines supply what is wanted without extra effort and, furthermore, the appearance of a DMU as part of an optimal basis ensures that it is efficient so that separate computations need not be made for these entities if that is all that is wanted.^{1/}

^{1/} Computer codes are available for effecting these computations. See [6]. New software by I. Ali and J. Stutz is also available from the Center for Cybernetic Studies at The University of Texas at Austin which detail the efficient facets observed.

It might be observed that the two inefficient DMU's that were accorded efficiency values of $h_0^* = 1$ have no such reference sets. This suggests that they have special properties which can be submitted to further analysis by means of the non-Archmidean formulations that we touched on earlier in the text.^{1/} We shall not turn aside to deal with that topic. Instead we shall simply accept this identification of H10 and H13 as a possible weakness of DEA in the comparisons we are making with other techniques since (as in this case) it can happen.

1/Note also that neither H10 nor H13 enter into the reference set for any other DMU.

Table 1

<u>Efficient DMU's</u>	<u>DEA Efficiency Rating (E)</u>	<u>Efficiency Reference Set</u>
H1	1.0	
H2	1.0	
H3	1.0	
H4	1.0	
H5	1.0	
H6	1.0	
H7	1.0	

<u>Inefficient DMU's</u>	<u>DEA Efficiency Rating (E)</u>	<u>Efficiency Reference Set</u>
H8	0.99	H4
H9	0.98	H1, H2, H6
H10	1.0	
H11	0.85	H4, H7
H12	0.99	H1, H4, H6
H13	1.0	
H14	0.99	H1, H4, H6
H15	0.87	H4, H6, H7

6. Cost Ratio Analysis

We now consider how a manager, e. g., in a rate setting commission for some state,^{1/} might determine which DMUs are more and less efficient when using ratios, a widely used form of analysis to evaluate financial and operating performance. In this example, all the inputs are jointly used by these DMUs to produce three outputs so that we cannot proceed as we might in the single output case. A number of different ratios might be developed to evaluate different sets of relationships such as FTEs/patient, FTEs/severe patient, FTEs/regular patient, FTEs/teaching output, bed days/patient, bed days/severe patient, etc. Such a set of ratios does not explicitly recognize the joint use of these inputs to produce these various outputs. In addition, for the set of ratios calculated, a DMU may be among the highest (least efficient) for certain ratios and lowest (most efficient) for other ratios. This leads to some ambiguity as to whether that DMU is efficient or inefficient and calls for some method of weighting or ordering the importance of the ratios to gain some overall assessment of efficiency such as was generated using DEA in Table 1.

Rather than address this issue directly, we will focus on a type of unit costing ratio analysis that is often applied to hospitals and other organizations to evaluate DMU performance. By design we can say that all 15 hospitals (DMUs) paid the same price per unit for each type of input and thus ignore possible difficulties which arise for a ratio analysis when this is not the case. That is, we can combine the inputs into dollar units without the confounding effect of differing input costs. Rather

^{1/} For instance, see [23] and [24].

than deal with all these outputs, the teaching output might be viewed as a by-product or secondary output and the patients might be viewed as a single output rather than segregate this into different categories of severity. This simplifying procedure is not wholly defensible from a cost accounting standpoint. Nevertheless, in the absence of any other way of combining and weighting the outputs, similar approaches have been used for hospitals as well as other types of DMUs (see for example [23]), and this is the way we shall proceed.

Table 2 column (A) reflects the average cost per patient for each DMU. This results in a ranking of hospitals reflected by the parenthesized number directly to the right of the average cost figure in Table 2. The lowest cost (most efficient) DMU is ranked 1 and highest cost (least efficient) DMU is ranked 13. This ranking erroneously classifies H13 (ranked 6) as more efficient than H3 (rank 7) and H6 (rank 9) and it classifies H9 as more efficient than H6. In addition, there is no objective means for determining the cutoff cost level to segregate efficient and inefficient units.

If the efficient relative costs of certain outputs are known, the outputs can be weighted to reflect a cost per weighted unit of output. In this case we know the efficient cost of a regular patient (\$130) and a severe patient (\$170) and the patient units can therefore be weighted to value each severe patient as the equivalent of $170/130 \sim 1.3$ regular patients. For example, H1 would have adjusted patient output units of 3000 regular patients + 2000×1.3 severe patients for an adjusted total of 5600 patients. Dividing this patient total into \$775,500, the total cost for H1 shown in Exhibit 3, results in \$138.48, the case mix adjusted average cost shown for H1 in column (B) of Table 2.

Table 2

Single Output Measures

<u>Hospital Efficient Units</u>	<u>Average Cost per Patient (A)</u>	<u>Case Mix Adjusted Average Cost per Patient (B)</u>	<u>Case Mix Adjusted Average Cost per Patient Segregated into High and Low Levels of Teaching Outputs</u>	
			<u>Low*</u> (C)	<u>High*</u> (D)
H1	\$155.10 (2)	\$138.48 (4)	\$138.48 (2)	
H2	163.32 (5)	138.40 (3)	138.40 (1)	
H3	168.32 (7)	142.65 (8)		\$142.65 (3)
H4	160.10 (4)	142.94 (9)	142.94 (6)	
H5	158.38 (3)	137.73 (2)		137.73 (2)
H6	170.15 (9)	140.12 (5)	140.12 (3)	
H7	142.60 (1)	135.81 (1)		135.81 (1)
<u>Inefficient Units</u>				
H8	176.95 (11)	157.99 (12)**		157.99 (6)
H9	168.32 (7)	142.64 (7)	142.64 (5)	
H10	169.69 (8)	161.61 (14)**		161.61 (7)
H11	170.33 (10)	153.10 (10)	153.10 (7)	
H12	178.33 (12)	155.07 (11)		155.07 (5)
H13	165.68 (6)	142.00 (6)	142.00 (4)	
H14	178.33 (12)	155.07 (11)		155.07 (5)
H15	179.74 (13)	160.48 (13)**	160.48 (8)	
Mean	167.02	146.94	144.77	149.42
Standard Deviation		8.82	7.36	9.66

* Low teaching outputs were 50 units and high teaching outputs were 100 units as per Exhibit 3, Col. 6.

**Hospitals more than one standard deviation over average cost.

The adjusted cost per patient is reflected in column (B) of Table 2 with the new ranking in parenthesis immediately to the right of the average cost per day. Even with this (normally not available) weighting of patients we continue to have a misranking with inefficient DMUs H9 and H13 being ranked as more efficient than H3 and H4. If we further segregate the 15 DMUs by the third output (teaching), as is sometimes done, and separate them based on those with high (100 units) versus low (50 units) teaching outputs, the ranking based on unit costs is reflected in columns C and D in Table 2. At this point, we have achieved an accurate ranking for the high teaching output hospitals but we still have not achieved an accurate ranking for the low ones. Because we have only two values for these outputs, at 50 and 100 "teaching units," we could distinguish high vs. low output hospitals fairly easily in the present case, but generally there will be many more values to consider with no objective guidance available for separating high from low teaching output values and the difficulty of distinguishing efficient from inefficient DMUs will then be compounded.

The problem of locating a point beyond which DMUs are considered inefficient is typically addressed by establishing a subjective cutoff value, even though there is no assurance, theoretical or otherwise, that the inefficient units will be accurately located through this process. For example, if the cutoff was set at one standard deviation above the mean adjusted cost per patient, only 3 DMUs (H8, H10 and H15) would be identified as inefficient as indicated in column (B) of Table 2.^{1/}

The DEA ratings in Table 1 do not lend themselves to rankings of the kind used in Table 2. As will be seen below, these efficiency measures

^{1/}At $0.6745\sigma = 5.95$, three more DMUs (H11, H12 and H14) would be added to this inefficient set. We record this as an additional possibility for improving this kind of identification even though most of the commonly used adjustments are in the direction of $k\sigma$, with $k > 1$.

are intended to supply estimates of excessive resource utilization relative to the Efficiency Reference Sets from which these ratings are derived. If, on the other hand, one uses the estimated resource savings as a basis and accords the same ranks to DMUs with equal efficiency ratings, a more informative set of ranks would be available from Table 1 than Table 2.^{1/} Whether ranked or not, however, Table 1 is more informative than Table 2 provided, of course, that the efficiency values exhibited in Table 1 are reasonably accurate.

7. Regression Analysis

In industries, including the "health industry," where the efficient input-output technology is not known with any real precision, regression analysis has been applied in order to gain "insights" into the production relationships that might underlie the observations that have been generated from past utilization of these processes. There are, of course, a variety of problems that are encountered when using traditional regression analyses to evaluate the efficiency of individual DMUs. One problem in most such studies is that one relatively smooth relation is posited to obtain the parameter estimates that are needed. Another problem is that the estimated parameter values are based on least squares estimates which

^{1/} In general one would also need to impute dollar magnitudes or other weights to the potential savings.

provide "mean" or "central tendency" values that reflect a mixture of efficient and inefficient behavior in the data set.^{1/} Thus, even if the posited functional forms are correct, the estimated regressions will only reflect efficient relationships if all units in the study are themselves efficient. Whatever reasons may be used to justify such assumptions in competitive industries, they are likely to be much weaker in not-for-profit settings such as education, health, and government.

Nevertheless such approaches have been extensively employed and so we now consider the extent to which regression analysis as it has been used, e. g., in health studies, might be employed to identify the inefficient units in the artificial data set. In the process we shall also locate other potential problems in the use of such analyses even when we can validly make the advantageous assumptions that all DMUs have the same technology and pay the same prices for all inputs.

One part of our analysis involves a simple linear (additive) regression model in which total cost was estimated as a function of the three outputs produced by each DMU. The results were as follows:

$$C = -95.300 + 152 y_1 + 182.4 y_2 + 1302 y_3$$

(8) (22.2) (767)

where C = Total cost per year (7)

y_1 = # of regular patients treated per year

y_2 = # of severe patients treated per year

y_3 = Training units provided in one year

^{1/} Recent literature has begun to supply a variety of means for addressing some of these problems when regression estimates for securing efficiency evaluations are wanted. They do not appear to be very satisfactory, however, and so we do not examine them here. See Banker, Charnes, Cooper and Maindaratta [2]. We confine ourselves only to those types of regressions which have been commonly (and widely) employed. See, e. g., [34].

The standard errors noted in the parentheses below each coefficient indicate high levels of statistical significance. The coefficient signs are positive, as required, and the relation between the y_1 and y_2 (for regular and severe patient) coefficients is in the correct (plausible) direction. A high R^2 value of 0.97 suggests a good fit with the observational data so, by standard reasoning, a high degree of cost variation is "explained" by these independent variables.^{1/}

The only apparent discrepancy is a fixed negative cost estimate of \$95,300. This value, which is not statistically significant, might cause the model to be questioned especially in cases involving hospitals with relatively small outputs. Hence another regression with its total cost intercept fixed at zero was calculated. We do not reproduce the results here, however, since (consistent with what has just been said) the resulting coefficient values did not differ greatly from those given in (7). Hence the latter might be used to estimate the incremental cost per unit of each output as in the second column of the following tabulation:

Output	Estimated Incremental Cost	Efficient Incremental Cost	% Deviation
y_1	\$ 152.	\$130	17.0
y_2	\$ 182.40	\$170	7.3
y_3	\$1302.	\$500	160.0

^{1/} The independent variables were found to have fairly low inter-correlations as follows:

$$r_{y_1y_2} = -0.37; \quad r_{y_1y_3} = -0.03; \quad r_{y_2y_3} = -0.08.$$

Focusing on the incremental costs in this manner bypasses the difficulties associated with a negative intercept value. It also corresponds to an assumption (not often stated explicitly) that the slope coefficients may still parallel the true incremental efficiency values, at least roughly, in a manner that corresponds to a shift of the regression plane up to the frontier without altering its slopes.^{1/} In the present case, we know the incremental costs for efficient operations and these are supplied in the third column. The estimates from the regression are high in every case. Only the estimate for y_2 (= severe patients) is even tolerable and the estimated cost for y_3 (= teaching) is very wide of the mark.

Another use of such regressions is to evaluate efficiencies as was done by Feldstein [18] in his now classic study of British hospitals. That is the actually observed outputs for each of H1 to H15 would be inserted in an expression like (7) and the resulting total cost would then be compared with the corresponding actual costs at this hospital.^{2/} The presence of a negative intercept value could be troublesome, however, and alternate forms of regression functions might then be explored.

^{1/} This method of parallel-shift treatment is explicitly incorporated in some of the "frontier estimation" methods that have recently been devised. See Forsund, Lovell and Schmidt [19].

^{2/} A variety of adjustments might be employed to allow for different hospital characteristics and patient mixes, etc. See Feldstein [18] for further discussion.

Another type of function that has been commonly employed in hospital studies, is the so-called Cobb-Douglas form. This form has the advantage of avoiding the possibility of negative intercepts and since, in the present data set, no zero outputs are present for any of the hospitals we can also avoid difficulties that are sometimes experienced from this quarter. Thus we now turn to such a Cobb-Douglas approach.

In logarithmic form our estimated relation obtained from the data of Exhibit 3 is

$$\ln C = 3.98 + .62 \ln y_1 + .57 \ln y_2 + .10 \ln y_3 \quad (8)$$

(.04) (.07) (.05)

which, in the usual Cobb-Douglas representation, becomes

$$C = 53.79 y_1^{0.62} y_2^{0.57} y_3^{0.10} \quad (9)$$

In this case the coefficients in (8) and hence the exponents in (9) all appear to be reasonable as well as significant. In sum, however, the exponent values (.62 + .57 + .20) exceed 1 which, being significant, means that evidence of decreasing returns to scale is present, or at least this possibility cannot be rejected. In our case this may reflect the complementary and substitution relations that are known to be present in some of the inputs.^{1/} The regression does not detect these relations in this form, however, and the fact that it results in a significant value (with $R^2 = 0.96$) could lead to erroneous recommendations with respect to decisions on the scale of operations.

^{1/} E. g., as reflected in $A^{-1} x = y$ when going from $x = Ay$, with A a matrix of positive constants as in (1). Thus, in general, A^{-1} will have negative as well as positive elements reflecting relations of complementarity as well as substitution among the various inputs used in producing these output combinations. See Sherman [29] for further discussion.

If we now consider DMUs as potentially inefficient when their actual total cost exceed the estimated total cost in (9), then efficient DMUs H2, H6, and H7 would be erroneously considered inefficient and inefficient DMUs H11, H12, H13, and H14 would be identified as efficient. These results together with the results of our preceding analysis are drawn together and presented in Table 3. In identifying which DMUs are efficient or inefficient, DEA has evidently done better than the others with the exception of the cost ratio approach when the latter is (a) adjusted for case mix and/or (b) identified with "low" and "high" levels of teaching outputs. There is, of course, a degree of arbitrariness present in these cost ratio efficiency and inefficiency characterizations that provide these favorable results for comparison with DEA. Furthermore the Case Mix adjustment procedure we used presupposes a knowledge of the efficient cost of operations and this is reflected in the results shown in both columns (B) and (C) in Table 3. Normally these costs will not be known and so we may count the apparently favorable results of these ratio analyses as proceeding from an assumed knowledge that will generally not be available. This knowledge is not required by DEA and hence we may regard it as being superior to the ratio analysis in these respects as well as in other respects that we shall begin to examine after first summarizing some of our other findings to this point as follows:

1. Ratio (cost) analysis and regression analysis required an arbitrary rule to determine which DMUs would be designated as inefficient. With ratio analysis, the mean might well have been lower or higher depending on whether there were more or fewer efficient units in the data set. Similarly, regression analysis might also have a lower or higher cost curve depending on the relative number of inefficient units.
2. Ratio analysis, as did regression analysis, required price data and other adjustments to address the multiple output and input situation while DEA could address this situation directly. In

addition, the ratios would be confounded if DMUs paid different prices for similar inputs. For example, a DMU that had very low prices might have a lower average cost that could obscure the presence of technical (production) inefficiencies. Regression analysis also assumed DMUs had the same costs/input, and different unit input costs would have shifted the cost function and could thereby also conceal inefficiencies.

3. Regression analysis results depended on the selection of an appropriate model or set of cost relationships and nothing in the data set suggested that either of the choices were not appropriate. DEA, however, required no such assumptions.

There are other points that can also be made as we move beyond mere classification into identifying the particular inputs where inefficiencies occur and estimating their amounts. This will be dealt with in the sections that follow.

Table 3

Comparison of DEA, ratio analysis, and linear regression approaches
ability to locate Inefficient DMU's

E = DMU rated as efficient
I = DMU rated as inefficient

	(A)	(B)	(C)	(D)
	DEA (1)	Ratio (2)	Case Mix Adjusted Average Cost/Patient (3)	Regression (4) (Cobb/Douglas)
Efficient DMU's	Results	Analysis		
H1	E	E	E	E
H2	E	E	E	I
H3	E	E	E	E
H4	E	E	E	E
H5	E	E	E	E
H6	E	E	E	I
H7	E	E	E	I
<u>Inefficient DMU's</u>				
H8	I	I	I	I
H9	I	E	E	I
H10	E	I	I	I
H11	I	E	I	E
H12	I	E	I	E
H13	E	E	E	E
H14	I	E	I	E
H15	I	I	I	I

(1) From table 1

(2) From table 2 column B - DMUs with cost/patient greater than one standard deviation above the mean used to identify inefficient DMUs.

(3) From Table 2 columns C and D with cost/patient greater than one standard deviation above the mean used to identify inefficient DMUs.

(4) Based on rule that DMUs with actual total cost greater than estimated total cost (based on the regression model) are inefficient.

8. Extensions

Perhaps the easiest approach to the topic of identifying the sources and estimating the amounts of inefficiency present in each DMU is to begin with a specific example. We therefore begin with H15 as an illustration of these kinds of additional uses of DEA. This hospital, which is inefficient, has already been discussed in association with H1 in Exhibit 3. We now approach it in a different manner as follows.

First consider the value of $h_0^* = 0.87$ in Table 1. Here we shall use this value to obtain the results shown in the column labelled "Intensity Adjusted Value" in Table 4. Because slack values also need to be considered in assessing efficiency we may refer to these h_0^* values as "intensity factors" and use them in the manner of the $h_0^* = 0.87$ value that is applied to each of the inputs in Table 4. The value which is then obtained in the case of H15 can then be compared with the corresponding value shown under the column labelled "True Efficiency Value". The latter are the values of the inputs actually needed for the outputs of H15 with efficient operations, as obtained from the efficient coefficient values provided in Exhibit 1. The

maximum discrepancy of $$(139,200-130,000) = \$9,200$ or, approximately, 7% occurs in the case of Supply \$. The other DEA estimates resulting from the intensity adjustment factor applied to the observed inputs are within 2% and 0.3%, respectively, of the true efficiency values.

TABLE 4
H15 INTENSITY ADJUSTMENT AND EFFICIENCY VALUE

Adjusted Input Values				Efficient Input Values			
	Observed Input Value	Intensity Adjustment Factor	Intensity Adjusted Value	Adjustments			True Efficiency Value
				Regular	Severe	Teach Units	
FTE:	26.5	x 0.87	= 23.055	.004 x 3,000 +	.005 x 2,000 +	.03 x 50	= 23.5
Bed Days:	47,370	x 0.87	= 41,211.9	(7 x 3,000 + 9	x 2,000)	÷ 0.95*	= 41,052
SUPPLY \$:	160,000	x 0.87	= 139,200	20 x 3,000 +	30 x 2,000 +	200 x 50	= 130,000

*0.95 = vacancy factor for efficient production. See assumption (a) in Exhibit 1.

Evidently our h_0^* value has operational significance in that it indicates "amounts" of inefficiency that are present. It thus differs from the index numbers and like approaches that are sometimes used for efficiency ratings. See, e. g., the index constructed by Feldstein [18] for use in the case of British hospitals.

As indicated earlier, the presence of slack in an optimal tableau is also to be considered a source of inefficiency, and these data, too, are available from the simplex tableaus. In particular, the slack value for Supplies in the optimal solution amounts to \$11,880 and 955 Bed Days of slack are also present. When these amounts are subtracted from the Intensity Adjusted Values in rows 3 and 2 of Table 4 new estimates for efficient inputs in these factors become \$127,313 and 40,257 BD, respectively. This greatly improves the efficiency estimate of the former along with some worsening of the latter. All estimates are now within about 2% of the true efficiency value.

It is not contended that DEA efficiency estimates will always be this close and, indeed, reference to Table 5 will show estimates that are very wide of the mark for H10 in at least 2 of the 3 pertinent input categories. On the other hand, even in this case the estimates are both better and more detailed than those obtained from the ratio and regression approaches discussed earlier in this article. Also, as was observed in our discussion of Table 1, there are strong reasons to suspect the $h_o^* = 1$ intensity values for H10 and H13. Elimination of these two hospitals still leaves H11 with errors in the range of 10-15% for three of the input estimates, while all of the other errors are in a range of about at 2% or less. Furthermore this record is considerably improved when the efficient hospitals, H1 to H7, are added to the list since in their case the estimates all have zero errors.

This seems to be a very creditable performance, at least compared to what the other approaches appear to offer for use on the data base we have erected. Further testing will also be required both on other data bases and in actual uses, of course, and improvements in the methodology and alternate modeling approaches and estimation methods will also need to be explored.

Methods by which such testing might be done will be discussed in the next section. We can then conclude this section by noting that still other uses of DEA are also possible. For instance, what we have been doing in this section amounts to projecting each DMU onto the relevant position of the efficiency surface in conformance with the methods prescribed in [13]. Further tradeoffs may then be effected by reference to the marginal rates of transformation and/or substitution via the optimal u_r^* and v_i^* values

which may be secured from the simplex tableaus. See (6). These values can provide guidance for augmenting or contracting the inputs and outputs of the corresponding DMU and, at the same time, provide controls and guidance on efficient uses by the managers of these DMUs.

These u_r^* and v_i^* values will represent estimates which, of course, may not be wholly accurate. The same is true of the similar uses of regression estimates but, in addition, such regression estimates can be expected to be very wide of the efficiency values--as should be clear from our earlier discussions. Indeed, as noted in [2], the estimates of such substitution and transformation rates generally continue to be very far from the true efficiency values even when the simple forms of regression functions used in the present article are replaced by more general and flexible forms and when the statistical methods used are specifically directed toward frontier efficiency estimates.

TABLE 5
ESTIMATED AND TRUE EFFICIENCY VALUES
H8 to H 15

HOSPITAL INPUTS		OBSERVED VALUE	INTENSITY ADJUSTED VALUE	SLACK	ESTIMATED EFFIC. VALUE	TRUE EFFIC. VALUE	% DIFF.
H8	FTE	25.0	24.75	--	24.75	25.0	1.0
	BD	49,475	48,980	8,425	40,555	41,053	1.2
	\$S	140,000	138,600	--	138,600	140,000	1.0
H9	FTE	24.5	24.01	--	24.01	24.5	2.0
	BD	43,160	42,297	--	42,297	43,158	1.9
	\$S	165,000	161,700	25,000	136,700	140,000	2.4
H10	FTE	77.0	77.0	--	77.0	53.0	45.0
	BD	92,630	92,630	--	92,630	92,632	0.0
	\$S	340,000	340,000	--	340,000	280,000	21.4
H11	FTE	44.5	37.8	5.1	32.7	36.5	10.4
	BD	65,260	55,471	--	55,471	65,263	15.0
	\$S	265,000	225,250	45,711	179,539	200,000	10.2
H12	FTE	30.0	29.7	--	29.7	30.0	1.0
	BD	60,000	59,400	9,476	49,924	50,526	1.2
	\$S	170,000	168,300	--	168,300	170,000	1.0
H13	FTE	43.5	43.5	--	43.5	43.5	0.0
	BD	81,110	81,110	--	81,110	76,842	5.6
	\$S	245,000	245,000	--	245,000	240,000	2.1
H14	FTE	30.0	29.7	--	29.7	30.0	1.0
	BD	60,000	59,400	9,476	49,924	50,526	1.2
	\$S	170,000	168,300	--	168,300	170,000	1.0
H15	FTE	26.5	23.06	--	23.06	23.5	1.9
	BD	47,370	41,212	955	40,256	41,053	1.9
	\$S	160,000	139,200	11,887	127,313	130,000	2.1

Note: H10 and H13 have intensity values of $h_0^* = 1$.

9. Conclusion

The really surprising result is not how well DEA performed on our manufactured data base, but rather the poor performance of the econometric-statistical models we employed. These models are representative of many analyses that have been employed in studies used to draw important policy conclusions. Two recent multi-million dollar studies of this kind that resulted in multi-volume reports with important findings for policy formation are: (1) U.S. Department of Health, Education and Welfare, PSRO: An Initial Evaluation of the Professional Standards Review Organization [in Health Care Delivery]^{1/} and (2) U.S. Office of Education, The Follow Through Planned Variation Experiment [for Education of Disadvantaged Children].^{2/}

The questions raised by our across-DMU regression results would seem to apply a fortiori to studies like these since in our case the design of the data base was favorable to assumptions such as a common technology and a common price structure across the DMUs. Assumptions like these are much less likely to be valid for regressions used in applied studies, such as the kinds we just cited.

It might be argued that it is unfair to level criticisms such as these at regression models designed to handle only one dependent variable at a time and using methods of estimation directed toward average rather than efficient behavior.^{3/} In the study [2], which we conducted with R. Banker and A. Maindiratta, however, both of these qualifications were accommodated.

^{1/} See [32]. See also [17] for further discussion and suggestions for alternative approaches.

^{2/} See [33]. See also [12] for further discussion and suggested alternative approaches.

^{3/} Note, however, the study by Feldstein [18] which was conducted in just this manner and numerous other studies of this type can also be cited. See also the study by Banker, Conrad and Strauss [4] which consisted of a DEA redo of a previously conducted econometric study of North Carolina hospitals (using a translog function) and arrived at drastically different conclusions on the presence of returns to scale, etc., which had been found not to be present in the original (econometric) study.

In that study, conducted in the same spirit as the one we are presently summarizing, a piecewise Cobb-Douglas function with one output as the dependent variable was used to represent a continuous technology with increasing and decreasing returns to scale in its various segments. Technical as well as scale inefficiencies were then introduced into randomly generated observations as a basis for comparing DEA with so-called flexible functional form approaches using translog regressions. DEA again performed very well but, perhaps even more importantly, the statistical-econometric approaches performed poorly--not only relative to DEA but also in a manner that was unsatisfactory per se--in both technical and scale efficiency identification and estimation. Moreover, the estimation methods employed for the regressions in this case were of the so-called "corrected least squares" varieties, as specifically designed for the purpose of locating and estimating efficiency frontiers. See [26] and [19].

One possible source of trouble, we think, lies not merely in the estimation methods but rather in an approach--the one that is commonly taught and employed--which tries to capture a great variety of behaviors in only relatively smooth and simple (e. g., unconstrained) functional forms. Attempts to meet these difficulties by weighted regressions, outlier analyses and similar approaches do not really deal with the problem in a sufficiently fundamental way, we think, and other alternatives need to begin to be considered.

The optimizations involved in these DEA and statistical approaches also need to be considered. Generally speaking the commonly employed statistical approaches optimize over all observations while DEA optimizes relative to each. Another way of stating this is to note that a complete DEA analysis will, in general, involve n optimizations, one for each

observation, while the usual statistical approach involves only one.

This implies that differences in testing for results and checking for possible inferences must also be expected. - Because it is directed toward individual observations, DEA is also directed to each DMU in a way which suggests this as a fundamental unit of test. That is, the inferences that are made about at least some of these DMUs can and should be tested by on-site observations in ways, and with results, that differ from testing statistical estimates for general types of class properties effected across all observations.

Having identified these differences and their possible separate avenues of application, testing and research, we can probably best close on a somewhat different note by indicating ways in which the two approaches might be joined together. One possibility is to use each approach, regression or ratio analysis and DEA, to check on or fortify the other.^{1/} Other possibilities exist, however, which might briefly be sketched as follows.

Aigner and Chu in [1], essayed a new approach to frontier estimation by means of what would now be called "goal programming"^{2/} with only one-sided deviations permitted so that, in general, the estimated production function (e. g., a Cobb-Douglas form) would lie on or above all of the observed output values. Confining all deviations to one side clearly does not exhaust the possibilities, however, and one may go on to prescribing proportions of the total deviations or even deviations for individual observations that must lie on one side or the other of an estimated frontier.

In a similar spirit, C. P. Timmer in [30] used "chance constrained programming" formulations and concepts to effect efficiency estimates.

^{1/} See [12] for further discussion on different conditions which might lead to one approach or the other in complementary fashion for policy guidance purposes

^{2/} This was originally referred to as "inequality constrained regressions." See [10] and [8]. Although not available at the time of the Aigner-Chu work [1] we would now add the further possibilities that are now available from the goal interval programming approaches described in [9].

Instead of utilizing the power of chance constrained programming, e. g. to deal with different proportions and even different probability distributions, constraint by constraint, Timmer proceeded in an entirely different direction and in the spirit of a "global" statistical analysis discarded "outlier" observations one after another until he achieved what he regarded as "stable" estimates. Notice, however, that this procedure is one which obliterates a great deal of information. In particular, in pursuit of one global (overall) property,^{1/} it discards efficient DMUs without even bothering to investigate them individually.

The approaches by Aigner and Chu [1] and by Timmer [30] that we have just described involve a use of inequality constrained optimizations, to be sure, but they otherwise proceeded in the spirit of classical statistical approaches. Something more may also be accomplished along these latter lines. For instance, one might use a discriminant-function or cluster-analytic approach to locate subsets of the original points which have different properties. Hopefully this could include clusters or discriminant subsets of efficient and inefficient points. Separate regressions fitted to these subsets might then yield improved ways of identifying inefficiencies and estimating their amounts.

We have not investigated the latter types of topics, as we shall do in future papers, for the simple reason that we sought to adhere as closely as possible to the kinds of approaches that have generally been used in the kinds of studies we have been considering. Notice that a use of the discriminant and/or cluster analysis approaches we have just described involves an estimation of more than one regression relation and more than one

^{1/} This is contrary to the spirit of individual observation investigation that we urged, above, and for which the kind of stability analysis provided in [11] is now available.

optimization. The other approaches of global programming and chance constrained programming varieties, as in Aigner and Chu [1] and Timmer [30], involve inequality constrained relations of a kind that are similar to the ones used in DEA. Thus, we conclude that there are additional avenues of possible relations between DEA and these other approaches that also invite exploration.

REFERENCES

- [1] Aigner, D. T. and S. F. Chu, "On the Estimation of the Industry Production Function," American Economic Review, Vol. 55, No. 4, September 1968, pp. 826-839.
- [2] Banker, R. D., A. Charnes, W. W. Cooper and A. Maindiratta "A Comparison of DEA and Translog Estimates of Production Frontiers Using Simulated Observations from a Known Technology," Research Report, Center for Cybernetic Studies, The University of Texas at Austin, October, 1982.
- [3] _____ "Estimation of Technical and Scale Efficiencies in Data Envelopment Analysis," Management Science, (forthcoming).
- [4] _____, R. Conrad, and R. Strauss, "An Application of Data Envelopment Analysis to A Study of Hospital Efficiencies and Scale Sizes," Working paper 81-9, Carnegie-Mellon University, December, 1981.
- [5] Bessent, A. and W. Bessent, "Determining the Comparative Efficiency of Schools Through Data Envelopment Analysis," Research Report CCS 361 (Austin, Texas: Center for Cybernetic Studies, The University of Texas at Austin, December 1979).
- [6] _____, J. Kennington and B. Regan, "An Application of Mathematical Programming to Assess Productivity in the Houston Independent School District," Management Science (forthcoming).
- [7] Charnes, A. and W. W. Cooper "An Efficiency Opening for Managerial Accounting in Not-for-Profit Entities" in H. P. Holzer, ed., Management Accounting 1980: Proceedings of the University of Illinois Management Accounting Symposium (Urbana: University of Illinois, 1980).
- [8] _____ "Goal Programming and Constrained Regression," Omega 3, No. 4, 1975, pp. 403-404.
- [9] _____ "Goal Programming and Multiple Objective Optimizations," European Journal of Operational Research 1 No. 1, 1977, pp. 39-54.
- [10] _____ and R. O. Ferguson, "Optimal Estimation of Executive Compensation by Linear Programming," Management Science 1 No. 2, 1955, pp. 138-151.
- [11] _____ A. Lewin, R. Morey and J. Rousseau, "An Approach to Positivity and Stability Analysis in DEA" Management Science (submitted).
- [12] _____ and E. Rhodes, "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through," Management Science, Vol. 27, No. 6, June 1981, pp. 668-697.
- [13] _____ "Measuring Efficiency of Decision Making Units," European Journal of Operational Research, Vol. 3, No. 4, July 1979, p. 339.

References (Continued)

- [14] _____ "Short Communication: Measuring Efficiency of Decision Making Units," European Journal of Operational Research, Vol. 3, No. 4, July 1979, p. 339.
- [15] _____, L. Seiford and J. Stutz, "A Multiplicative Model for Efficiency Analysis," Socio-Economic Planning Sciences (forthcoming).
- [16] _____ "Invariant Multiplicative Efficiency and Piecewise Cobb-Douglas Envelopments," Operations Research Letters (forthcoming).
- [17] Churchill, N. C., W. W. Cooper and V. Govindarajan, "Effects of Audits on the Behavior of Medical Professionals Under the Bennett Amendment," Auditing: A Journal of Practice and Theory Vol. 1, No. 2, Winter 1982.
- [18] Feldstein, M., Economic Analysis for Health Service Efficiency (Amsterdam: North-Holland Press 1968).
- [19] Forsund, F. R., C. A. K. Lovell, and P. Schmidt, "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement," Journal of Econometrics, Vol. 13, No. 1, 1980, pp. 5-26.
- [20] 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., 1977.
- [21] Lewin, A. Y. and R. C. Morey, "Evaluating the Administrative Efficiency of Courts," Omega (forthcoming).
- [22] _____ "Studies to Validate and Extend the Data Envelopment Analysis Method for Measuring Relative Efficiency of Decision Making Units," Durham, N. C. Duke University, Fuqua School of Business, 1980.
- [23] Massachusetts Rate Setting Commission, Development of a Budget Approval Program for Massachusetts Hospitals, 1976-1978, Final Report, 12/1/78, HCFA contract #600-76-0174, Vol. 2, Chapter 8 (Washington, D. C. HCFA, 1978).
- [24] _____ Hospital Bureau, "Grouping of Hospitals for Purposes of Determining Reasonable Cost Levels," Document No. 114.1 CMR, Rate Setting Commission, 1980.
- [25] Rhodes, E., Data Envelopment Analysis and Related Approaches for Measuring the Efficiency of Decision Making Units with an Application to Program Follow Through in U. S. Education Ph.D. Thesis (Pittsburgh: Carnegie-Mellon University, School of Urban and Public Affairs, 1978).

References (Continued)

- [26] Richmond, J., "Estimating the Efficiency of Production," International Economic Review, 1974, pp. 515-521.
- [27] Sato, D., Production Functions and Aggregation (Amsterdam: North Holland Publishing Co., 1975).
- [28] Schinnar, A. P. "Frameworks for Social Accounting and Monitoring of Invariance, Efficiency and Heterogeneity," Fels Discussion Paper 147 (Philadelphia: University of Pennsylvania, School of Public and Urban Policy, 1980).
- [29] Sherman, H. D., Measurement of Hospital Technical Efficiency: A Comparative Evaluation of Data Envelopment Analysis and Locating Inefficiency in Health Care Organizations, unpublished Doctoral Dissertation, Boston; Harvard University Graduate School of Business Administration, 1981.
- [30] Timmer, C. P. "Using a Probabilistic Frontier Production Function to Measure Technical Efficiency," Journal of Political Economy LXXIX No. 4., July-Aug., 1971, pp. 776-794.
- [31] U.S. Congressional Budget Office Background Paper, The Effect of PSROs on Health Care Costs: Current Findings and Future Evaluations (Washington: U.S. Government Printing Office, June, 1979).
- [32] U.S. Department of Health, Education and Welfare, PSRO: An Initial Evaluation of the Professional Standards Review Organization Vol. 1, Executive Summary, Washington, 1978, National Technical Information Service II PB 278154/AS, HEW Report OPEL 77-12, Feb., 1978.
- [33] U.S. Office of Education, The Follow Through Planned Variation Experiment, Vols. IIA, IIB, IIC. Washington, 1977.
- [34] U.S. General Accounting Office, "Problems with Evaluating the Cost Effectiveness of Professional Standards Review Organizations," Report by the Comptroller General of the United States HRD-79-52 (Washington: U.S. General Accounting Office, July, 1979).
- [35] Zaretsky, H. W. "The Effect of Patient Mix and Service Mix on Hospital Costs and Productivity," Topics in Health Care, A. Herkimer, Ed., Vol. 4, No. 2, Winter 1977, pp. 63-82.

APPENDIX

Exhibit 1

Appendix -

Structural Model (Efficient Hospital Operations)

Efficient Input-Output and Cost Relationships Assumed in the

Hospital Production Model to Create
the Simulated Data Base in Exhibit 2

Amount of each Input required to Efficiently produce one unit of output.			Efficient Costs of Outputs	
Outputs	Full Time Equivalents of Labor (FTE's)	Bed days available	Supply \$'s	
Regular Patient Care	.004 FTE/per patient	7 beddays/per patient	\$20 supplies/per patient	\$130/regular (1) patient
Severe Patient Care	.005 FTE/per patient	9 beddays/per patient	\$30 supplies/per patient	\$170/severe (2) patient
Training outputs	.03 FTE/per training unit	-	\$200. supplies/per training unit	\$500/training (3) unit

Input \$ Costs for all Hospitals

FTE's	Beddays	Supply's
\$10,000/FTE	\$10/Bed day	\$1/supply unit

Other Assumptions:

- Vacancy Rate - Efficient hospitals will have 5% of total beds vacant during the year (available for emergencies).
 - There are no regional cost differences for bed days, supplies, and FTE's and that the mix of FTE's and supplies are similar between hospitals.
 - Cost of unused bed days = \$10/bed day.
- Cost/regular patient = (.004 FTE/patient) (\$10,000/FTE) + (7 bed days/patient) (\$10/bed day) + (\$20 supplies/patient) = \$130/patient
 - Cost/severe patient = (.005 FTE/patient) (\$10,000/FTE) + (9 bed days/patient) (\$10/bed day) + (\$30 supplies/patient) = \$170/patient
 - Cost/training unit = (.03 FTE/training unit) (\$10,000/FTE) + (\$200 supplies/training unit) = \$500/patient

Exhibit 2

Appendix

Constructed Data Base

□ - Inefficient use of inputs compared to a structural model of efficient input-output relationship described in Ex. 1

Inputs				Outputs				Actual Inputs Used									
FFE	Bed Days	Supply \$'s	Total Pac's	% Sev. Pac's	Teach. Units	Reg. Pat's	Sev. Pat's	FTE		Bed days		Supply \$		Training		Vacancy Rate	
								Reg. Pat's	Sev. Pat's	Reg. Pat's	Sev. Pat's	Reg. Pat's	Sev. Pat's	Supply \$'s	FTE		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	
H1	23.5	41050	\$130,000	5000	40	50	3000	2000	.004	.005	7	9	20	30	200	.03	.05
H2	24.5	43160	140,000	5000	60	50	2000	3000	.004	.005	7	9	20	30	200	.03	.05
H3	26.0	43160	150,000	5000	60	100	2000	3000	.004	.005	7	9	20	30	200	.03	.05
H4	25.0	41050	140,000	5000	40	100	3000	2000	.004	.005	7	9	20	30	200	.03	.05
H5	28.5	50530	160,000	6000	50	50	3000	3000	.004	.005	7	9	20	30	200	.03	.05
H6	36.0	62105	210,000	7000	71	100	2000	5000	.004	.005	7	9	20	30	200	.03	.05
H7	51.5	92630	270,000	12000	17	50	10000	2000	.004	.005	7	9	20	30	200	.03	.05
H8	25.0	49475	140,000	5000	14	100	3000	2000	.004	.005	9	10	20	30	200	.03	.05
H9	24.5	43160	165,000	5000	60	50	2000	3000	.004	.005	7	9	20	30	200	.03	.05
H10	77.0	92630	340,000	12000	17	100	10000	2000	.006	.007	7	9	20	30	200	.03	.05
H11	44.5	65260	265,000	8000	38	50	5000	3000	.006	.007	7	9	20	30	200	.03	.05
H12	30.0	60000	170,000	6000	50	100	3000	3000	.004	.005	7	9	20	30	200	.03	.05
H13	43.5	81110	245,000	9000	56	50	4000	5000	.004	.005	7	9	20	30	200	.03	.05
H14	30.0	60000	170,000	6000	50	100	3000	3000	.004	.005	9	9	20	30	200	.03	.10
H15	26.5	47370	160,000	5000	40	50	3000	2000	.005	.005	9	9	20	30	200	.03	.05

* Total Patients = Col. 7 + Col. 8
 ** % Severe = Col. 8 + Col. 4

Note: Outputs (Col.s 4 - 8) and input usage (Col.s 9 - 17) are arbitrarily assigned so that H1 - H7 are efficient based on the production model in Ex. 1 and H8 - H15 are inefficient based on the same model. Inputs (Col. 1, 2, 3) are derived from Col.s 4 - 17 as indicated in Ex. 3

Exhibit 3

Appendix

Example of Construction of Data Base for Hospitals H1 (efficient) and H15 (inefficient)

Outputs	Efficient		Inefficient		Difference Between H1 and H15	Inputs Required for Each Unit of the Related Output				Vacancy Rate during the Year	
	H1	H15	H1	H15		FTE/yr.	Beds days/yr.	Supply \$/yr.		H1	H15
Y_1 Regular patients/yr.	3000	3000	3000	3000	-	.004	7	\$20	4	52	52
Y_2 Severe patients/yr.	2000	2000	2000	2000	-	.005	9	30	9		
Y_3 Teaching units/yr.	50	50	50	50	-	.03	-	200	-		

(Outputs are identical for H1 and H15)

Total Inputs Required			
FTE	23.5 ⁽¹⁾	26.5 ⁽²⁾	3
Beds days	41,050 ⁽³⁾	47,370 ⁽⁴⁾	6,320
Supplies	\$130,000 ⁽⁵⁾	\$160,000 ⁽⁶⁾	\$30,000
Total Cost	\$775,500 ⁽⁷⁾	\$898,700 ⁽⁸⁾	\$123,200

* Vacancy rate reflects the % of vacant beds that exist. Beds available to exceed the number used by 1/.95 in efficient hospitals (5% are vacant). If 7 beds are used, 7/.95 must be available on average during the year.

- (1) H1-FTE's = (3000 Reg. Pat.)(.004) + (2000 Sev. Pat.)(.005) + (50 Teach. Units)(.03) = 23.5
- (2) H15-FTE's = (3000 Reg. Pat.)(.005) + (2000 Sev. Pat.)(.005) + (50 Teach. Units)(.03) = 26.5
- (3) H1-Beds days = [(3000 Reg. Pat.)(7) + (2000 Sev. Pat.)(9)] + (.95 Vacancy factor) = 39,000 + .95 = 41,050
- (4) H15-Beds days = [(3000 Reg. Pat.)(7) + (2000 Sev. Pat.)(9)] + (.95 Vacancy factor) = 45,000 + .95 = 47,368
- (5) H1-Supply \$'s = (3000 Reg. Pat.)(20) + (2000 Sev. Pat.)(30) + (50 Teach. Units)(200) = \$130,000
- (6) H15-Supply \$'s = (3000 Reg. Pat.)(20) + (2000 Sev. Pat.)(30) + (50 Teach. Units)(200) = \$160,000
- (7) H1-Total Cost = (23.5 FTE)(\$30,000/FTE) + (41,050 Bed days x \$10/bed day) + \$130,000 Supplies = \$775,500
- (8) H15-Total Cost = (26.5 FTE)(\$30,000/FTE) + (47,368 Bed days x \$10/bed day) + \$160,000 Supplies = \$898,700