Journal of Econometrics 30 (1985) 91–107. North-Holland

FOUNDATIONS OF DATA ENVELOPMENT ANALYSIS FOR PARETO–KOOPMANS EFFICIENT EMPIRICAL PRODUCTION FUNCTIONS*

A. CHARNES, W.W. COOPER, B. GOLANY and L. SEIFORD

The University of Texas, Austin, TX 78712, USA

J. STUTZ

University of Miami, Coral Gables, FL 00000, USA

The construction and analysis of Pareto-efficient frontier production functions by a new Data Envelopment Analysis method is presented in the context of new theoretical characterizations of the inherent structure and capabilities of such empirical production functions. Contrasts and connections with other developments, including solutions of some remaining problems, are made re aspects such as informatics, economies of scale, isotonicity and non-concavity, discretionary and non-discretionary inputs, piecewise linearity, partial derivatives and Cobb–Douglas properties of the functions. Non-Archimedean constructs are *not* required.

1. Introduction

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 Shephard (1953, 1970), under restrictions on the mathematical structure of production possibility sets and cost relations, developed an elegant 'transform' theory between production aspects and cost aspects as in Charnes, Cooper and Schinnar (1982). 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 like Afriat (1972), Aigner, Lovell and Schmidt (1977), and Försund and Hjalmarsson (1979).[1] These efforts were almost exclusively for single-output functions.

Farrell (1957), partly responding to the inadequacies of separate indices of labor productivity, capital productivity, etc., undertook what he referred to as

*This research was partly supported by NSF Grant SES-8408134 and USARI Contract MDA903-83-K-0312 with the Center for Cybernetic Studies, The University of Texas at Austin. Reproduction in whole or in part is permitted for any purpose of the Unisted States Government.

[1] The need for some method of estimating frontier functions had been pointed out as early as 1935 by R. Frisch (1935) in his statistical study of chocolate production in France.

an 'activity analysis' approach[2] that could deal more adequately with the problem. Possibly because of the limitations of the elaborate matrix inversion routines he was employing, Farrell confined his numerical examples and discussion to single-output situations, although he did formulate the multiple-output case. If anything, the need for being able to deal directly with multiple outputs has grown ever more urgent with the continuing growth of the not-for-profit and service sectors in the United States and other 'western type' economies. As we have elsewhere observed,[3] one can otherwise do almost nothing with important parts of not-for-profit and governmental activities without becoming involved in arbitrary weighting schemes. Even attempts at measuring and evaluating 'total factor productivity' will fall short of what is required since this is inherently a single-output concept. Farrell was right, therefore, in re-orienting the direction of development toward the various types of efficiencies that he referred to as 'technical', 'scale', and 'allocative efficiencies'.

Building on the individual firm or country (= Decision Making Unit)[4] evaluations of Farrell and the engineering ratio idea of efficiency measure for a single-input, single-output efficiency analysis in its managerial aspects and its constructible extensions to multi-input, multi-output situations was initiated by Charnes, Cooper and Rhodes (1978, 1981). Subsequent extensions and elaborations by the former pair with other students and colleagues [Charnes, Cooper, Lewin, Morey and Rousseau (forthcoming), Charnes, Cooper, Seiford and Stutz (1982, 1983), and Charnes, Cooper and Sherman (forthcoming)] were made with more detailed attention to classical economic aspects and deeper analysis of the production function side of the mathematical duality structure and Data Envelopment Analysis first presented in the original CCR work. The CCR ratio measures and the variants of Farrell, Shephard, Färe et al. on the dual linear programming side 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 work by others. 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

[2] In the sense of the original work by T.C. Koopmans that is described in Chapter IX of Charnes and Cooper (1980) where it was also accorded an explicit linear programming formulation and interpretation.

[3] See Banker, Bowlin, Charnes and Cooper (in process).

[4] This term was introduced in Charnes, Cooper and Rhodes (1978) and contracted to DMU because it was recognized that something like it was needed to refer to public sector organizations.

Analysis method and informatics which permit a constructive development of an empirical production function and its partial derivatives without loss of efficiency analysis or use of non-Archimedean field extensions.

2. 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 'output'. 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 'DMUs', 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 observed values are to be non-negative (sometimes positive) numbers.

We shall employ the notation X_j, Y_j for respectively the observed vectors of the inputs and outputs of the jth DMU. By X and Y we shall mean the *matrices* whose n column vectors are respectively those of the n DMUs. We use x, y, λ and s to denote column vectors, also with e the column vector of ones. We use 'T' as a superscript to denote the transpose, e.g. e^{T} is the *row* vector of ones. Point sets are denoted by (capital) script letters. The inequality symbol between vectors means that inequality holds for each component.

3. A hypograph empirical production possibility set

Given the (empirical) points (X_j, Y_j), $j = 1, \ldots, n$, 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' $\mathscr{P}_E$ to be the convex hull of these points, i.e.,

$$\mathscr{P}_E \triangleq \left\{ (x, y) : x = \sum_{j=1}^{n} X_j \mu_j ,\ y = \sum_{j=1}^{n} Y_j \mu_j , \forall \mu_j \geq 0, \sum_j \mu_m = 1 \right\}. \tag{3.1}$$

We extend it to our 'empirical production possibility set' $\mathscr{Q}_E$ by adding to $\mathscr{P}_E$ all points with inputs in $\mathscr{P}_E$ and outputs not greater than some output in $\mathscr{P}_E$, i.e.,

$$\mathscr{Q}_E \triangleq \{ (x, y) : x = \bar{x},\ y \leq \bar{y} \text{ for some } (\bar{x}, \bar{y}) \in \mathscr{P}_E \}. \tag{3.2}$$

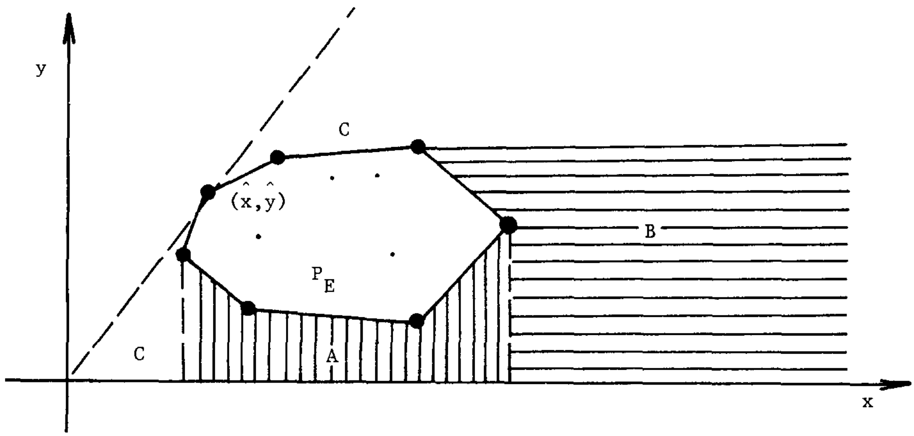

Fig. 1

Note that $\mathcal{Q}_E$ is contained in (e.g., is smaller than) *every* production possibility set heretofore employed, i.e., those studies by Farrell (1957), Shephard (1970), Banker, Charnes and Cooper (1984), Fare and Lovell (1978), etc. We also use fewer axioms than the others, including even Banker (1984) who, up to this point in *journal* publication, had used the most parsimonious axiom system to characterize the production possibility set of DEA.[5] The Farrell, Shephard, Färe sets are (truncated) cones; the BCC set (when not also a cone) adds to $\mathcal{Q}_E$ the set

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

These relations may be visualized in the schematic plot of fig. 1, where $\mathcal{Q}_E = \mathcal{P}_E \cup \mathcal{A}$, the BCC set is $\mathcal{Q}_E \cup \mathcal{B}$, and the Farrell, Shephard, Färe set is $\mathcal{Q}_E \cup \mathcal{B} \cup \mathcal{C}$.

Let $\mathcal{P}_E^t$, $\mathcal{Q}_E^t$ denote the sets corresponding to $\mathcal{P}_E$ and $\mathcal{Q}_E$ when only the output y_t is the ordinate. Evidently a frontier function $f_t(x)$ is determined by

$$f_t(x) = \max y_t \quad \text{for} \quad (x, y_t) \in \mathcal{Q}_E^t. \tag{3.3}$$

Then:

Theorem 1. *$\mathcal{Q}_E^t$ is the hypograph of $f_t(x)$ over $\{x: (x, y) \in \mathcal{Q}_E\}$.*

[5] Our usage antedates his since we used this in the Ph.D. and other work with Rhodes (the 'R' of CCR).

Proof. The hypograph H_ι of $f_\iota(x)$ is the set

$$H_\iota \triangleq \{(x, y_\iota): y_\iota \le f_\iota(x), (x, y) \in \mathcal{Q}_E\}. \quad \blacksquare$$

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

Theorem 2. $f_\iota(x)$ is a concave, piecewise linear function on $\mathcal{D}_E$.

Proof. A necessary and sufficient condition for $f_\iota(x)$ to be concave is that its hypograph is a convex set [cf. Rockafellar (1970) or Fenchel (1953)]. The piecewise linearity also follows from the construction of $\mathcal{Q}_E$ by all convex combinations of the empirical points (X_j, Y_j), $j = 1, \ldots, n$. $\blacksquare$

We further observe explicitly that *no use* whatever has been made of *non-negativity* of input and output values in the sets, functions or proof of Theorems 1 and 2. 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 2 shows, therefore, that any *empirical* (maximum) frontier function is the 'concave cap' function of its graph.

4. The empirical Pareto-efficient production function

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

$$g_k(x) \le g_k(x^*), \qquad k = 1, \ldots, K, \tag{4.1}$$

with at least one strict inequality. Charnes and Cooper (1962, ch. IX) showed that x^* is Pareto-efficient iff x^* is an optimal solution to the mathematical (goal) program[6]

$$\min_x \sum_{k=1}^{K} g_k(x) \quad \text{subject to} \quad g_k(x) \le g_k(x^*), \qquad k = 1, \ldots, K. \tag{4.2}$$

This was employed by Ben-Israel, Ben-Tal and Charnes (1977) to develop the

[6]For a formal definition of goal programming and some of its history, see Charnes and Cooper (1977).

currently strongest necessary and sufficient conditions for a Pareto-minimum in convex programming.

Utilizing (4.2) we can now define and construct, im(or ex-)plicitly, the Pareto-efficient empirical (frontier) production function. Because of Koopmans' work in this area [see Charnes and Cooper (1961)], we shall use interchangeably the designations 'Pareto-efficient' and 'Pareto–Koopmans efficient', etc. in this paper. Other usages of (4.2) to generalizations such as the 'functional efficiency' of Charnes and Cooper (1961) will not be developed here.

First, by (4.2), the Pareto-efficient points among our n empirical points can be determined. The empirical Pareto-efficient function is then defined on the convex hull of their inputs by convex combinations of the 'output' values. Note that the convex hull of the Pareto-efficient points might not include all of $\mathscr{P}_E$ since only the doubled line portion of the frontier corresponds to Pareto-efficient points.

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.,

$$\begin{aligned} -g_k(x) &\triangleq y_k, \qquad 1 \leq k \leq s, \\ &\triangleq -x_i, \quad k = s+i, \quad i = 1, \dots, m, \quad \text{for } (x, y) \in \mathscr{Q}_E, \end{aligned} \tag{4.3}$$

For the optimization in (4.2) we clearly need only consider $(x, y) \in \mathscr{P}_E$ rather than $\mathscr{Q}_E$. Thus the constraint inequalities in (4.2) are for a test point (x^*, y^*):

$$y \geq y^*, \qquad x \leq x^*, \tag{4.4}$$

and we have, since these are the envelopment constraints of Data Envelopment Analysis for an observed input vector x^* and corresponding output vector y^*:

Theorem 3. The envelopment constraints of Data Envelopment Analysis in production analysis are the Charnes–Cooper constraints of (4.2) for testing Pareto–Koopmans efficiency of an empirical production point.

In no way is what we call 'Data Envelopment Analysis' restricted to linear constant returns to scale functions or to truncated cone domains. Evidently via (4.2), Data Envelopment Analysis applies to much more general convex 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–Koopmans efficiency, the Charnes–Cooper test of (4.2) becomes

$$\min_{\lambda, s^+, s^-} \; -e^{\mathrm{T}} Y \lambda + e^{\mathrm{T}} X \lambda,$$

subject to

$$Y\lambda - s^+ = Y_0, \quad -X\lambda - s^- = -X_0, \quad e^{\mathrm{T}}\lambda = 1, \tag{4.5}$$

$$\lambda, s^+, s^- \geq 0,$$

where

$$X \triangleq [X_1, \ldots, X_n], \qquad Y \triangleq [Y_1, \ldots, Y_n].$$

Since $-e^{\mathrm{T}}(Y\lambda - Y_0) + e^{\mathrm{T}}(X\lambda - X_0)$ 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

$$\min_{\lambda, s^+, s^-} \; -e^{\mathrm{T}}s^+ - e^{\mathrm{T}}s^-,$$

subject to

$$Y\lambda - s^+ = Y_0, \quad -X\lambda - s^- = -X_0, \quad e^{\mathrm{T}}\lambda = 1, \tag{4.6}$$

$$\lambda, s^+, s^- \geq 0.$$

Here, $\min -e^{\mathrm{T}}s^+ - e^{\mathrm{T}}s^- = -e^{\mathrm{T}}s^{+*} - e^{\mathrm{T}}s^{-*} = 0$ if and only if Y_0, X_0 is Pareto–Koopmans efficient.[7] This is the new DEA form for obtaining and characterizing the production possibility set $\mathcal{Q}_E$ via $\mathcal{P}_E$. The linear program (4.6) maximizes the l_1-distance of a point in $\mathcal{P}_E$ to (X_0, Y_0). We solve (4.6) for all n DMUs considered as (X_0, Y_0). From these we get the efficient and non-efficient points and can construct as desired the points of $\mathcal{P}_E$ by convex combinations. As we shall see later, other variations of $\mathcal{Q}_E$ can be accommodated easily by simple modifications of or additions to the constraints on λ. Its informatics and software, as developed by Ali and Stutz of the Center for Cybernetic Studies of The University of Texas at Austin, involve only minor modification from that of Charnes, Cooper, Seiford and Stutz (1983).

5. Efficiency analysis

As mentioned, managerial and program comparison aspects of efficiency analysis were initiated in Charnes, Cooper and Rhodes (1978, 1981) and Charnes and Cooper (1980), 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 'DMUs' (sometimes called 'productive' units or 'response' units). The multi-input, multi-output situations were reduced to 'virtual' single-input, single-output ones through use of virtual multipliers and

[7]See Charnes and Cooper (1961, ch. IX).

sums. Explicitly, the CCR ratio measure of efficiency of the DMU designated '0' is given by the non-linear, non-convex, non-Archimedean fractional program [see Charnes, Cooper, Lewin, Morey and Rousseau (forthcoming)]

$$\max_{\eta,\xi} \eta^{\mathrm{T}} Y_0/\xi^{\mathrm{T}} X_0,$$

subject to

$$\eta^{\mathrm{T}} Y_j/\xi^{\mathrm{T}} X_j \leq 1, \quad -\eta^{\mathrm{T}}/\xi^{\mathrm{T}} X_0 \leq -\varepsilon e^{\mathrm{T}}, \quad -\xi^{\mathrm{T}}/\xi^{\mathrm{T}} X_0 \leq -\varepsilon e^{\mathrm{T}}, \tag{5.1}$$

for $j=1,\ldots,n$, 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 η^{T}, m entries for ξ^{T}. (X_0, Y_0) is one of the n input–output pairs.

Employing the Charnes–Cooper transformation of fractional programming[8]

$$\mu^{\mathrm{T}} \triangleq \eta^{\mathrm{T}}/\xi^{\mathrm{T}} X_0, \quad \nu^{\mathrm{T}} \triangleq \xi^{\mathrm{T}}/\xi^{\mathrm{T}} X_0, \quad \nu^{\mathrm{T}} X_0 = 1, \tag{5.2}$$

we obtain the dual non-Archimedean linear programs

$$\max_{\eta,\nu} \mu^{\mathrm{T}} Y_0,$$

subject to

$$\nu^{\mathrm{T}} X_0 = 1, \quad \mu^{\mathrm{T}} Y - \nu^{\mathrm{T}} X \leq 0, \quad -\mu^{\mathrm{T}} \leq -\varepsilon e^{\mathrm{T}}, \quad -\nu^{\mathrm{T}} \leq -\varepsilon e^{\mathrm{T}}, \tag{5.3}$$

and

$$\min_{\theta,\lambda,s^+,s^-} \theta - \varepsilon e^{\mathrm{T}} s^+ - \varepsilon e^{\mathrm{T}} s^-,$$

subject to

$$Y\lambda - s^+ = Y_0, \quad \theta X_0 - X\lambda - s^- = 0,$$

$$\lambda, s^+, s^- \geq 0,$$

where

$$X \triangleq [X_1,\ldots,X_n], \qquad Y \triangleq [Y_1,\ldots,Y_n].$$

The first problem is associated with the origin of the term 'Data Envelopment Analysis' since the minimization (a) envelops the output vector Y_0 from above and (b) envelops the input vector X_0 from below via the minimizing choice of the scalar value of the intensity $\theta^* = \min\theta$. The second problem is said to be in efficiency analysis form with the maximization oriented toward the choice of μ^{T} and ν^{T} (called *virtual* multipliers or transformation rates)

[8]See Charnes and Cooper (1962) and Schaible (1974).

which produces the greatest rate of virtual output per unit virtual input allowed by the first constraint together with the requirements (a) virtual output cannot exceed virtual input and (b) all virtual transformation rates must be positive.

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

$$\theta - \varepsilon[e^{\mathrm{T}}Y\lambda - e^{\mathrm{T}}X\lambda] \tag{5.4}$$

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), $j = 1, \ldots, n$, with pre-emption on the intensity θ of input X_0. As mentioned above, 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 Charnes and Cooper (1961), 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 fig. 1.

If a DMU is inefficient, the optimal $\lambda_j^* > 0$ in its DEA problem (= Charnes–Cooper test) designate *efficient* DMUs, as do alternate optima. Thus, a 'proper' subset of the efficient DMUs 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 is associated. The transformation,

$$X_0 \to \theta^* X_0 - s^{*-}, \qquad Y_0 \to Y_0 + s^{*+}, \tag{5.5}$$

where the asterisk designates optimality, projects DMU_0, i.e., (X_0, Y_0), onto its efficiency facet.

This projection was introduced by Charnes, Cooper and Rhodes (1981) to correct for differences in managerial ability in order to distinguish between 'program' and 'managerial efficiency' 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.[9] Thus, although the relative efficiency *measure* of an inefficient

[9]The analysis by Bessent, Bessent, Charnes, Cooper and Thorogood (1983) shows how one might take account of the possible effects on other DMUs when one or more of the efficient DMUs is altered.

DMU will involve the infinitesimal ε, *non*-infinitesimal changes for improvement *are* suggested.

Both Farrell and Shephard knew that ratio measures required adjustments to correctly exhibit inefficiency of the second DMU in examples like the following two-input, one-output, two-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 was introduced 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.[10]

Our new Pareto-efficient DEA method like in Charnes, Cooper, Seiford and Stutz (1983) associates facets with non-optimal (= non-Pareto-efficient) DMUs. Clearly, by the Charnes–Cooper test, DMU_0 is Pareto-efficient (Pareto-efficient) DMUs. Clearly, by the Charnes–Cooper test, DMU_0 is Pareto-efficient (Pareto-optimal) iff $-e^{\mathrm{T}}s^{*+}-e^{\mathrm{T}}s^{*-}=0$, i.e., iff the l_1-distance from (X_0, Y_0) to the farthest 'northwesterly' $\mathscr{Q}_e$ point is zero. The CCR efficient DMUs are also among the new Pareto-efficient DMUs. Projection of a non-optimal DMU onto its Pareto-efficient facet is rendered by

$$X_0 \to X_0 - s^{*-}, \qquad Y_0 \to Y_0 + s^{*+}. \tag{5.6}$$

To achieve a convenient efficiency measure, we modify the functional by multiplying it by a $\delta > 0$ and dividing the s^+ and s^- by respectively the entries in Y_0 and X_0, e.g.,

$$-\delta e^{\mathrm{T}}D^{-1}(Y_0)s^{+} - \delta e^{\mathrm{T}}D^{-1}(X_0)s^{-}, \tag{5.7}$$

where $D^{-1}(Y_0)$ and $D^{-1}(X_0)$ are diagonal matrices whose diagonals are the reciprocals of the entries in Y_0 and X_0 respectively. This achieves a units-invariant measure which may be thought of as the *logarithm* of the efficiency measure. A $\delta = 10/(m+s)$ will yield a logarithm between 0 and -10. This measure might then be called the 'efficiency pH' analogy with the pH of chemistry.

Our new measure relates to the units-invariant multiplicative measure of Charnes, Cooper, Seiford and Stutz (1983) which is necessary and sufficient

[10] See the discussion in Charnes and Cooper (1961, p. 756).

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 X_j, Y_j which we employ in the multiplicative formulation.

6. Informatics and function properties

6.1. Partial derivatives

The guidance provided by the Charnes–Cooper–Rhodes, Banker–Charnes–Cooper, Charnes–Cooper–Seiford–Stutz 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 Charnes–Cooper for Pareto-efficiency, the following constructive method can be employed.

On reaching a non-Pareto-efficient point, our software discovers all the optimal observed points in its facet, hence, implicitly, all the convex combinations which form the facet. Since the Pareto-efficient 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.

Let

$$F(x_1, \dots, x_m, y_1, \dots, y_s) = 0 \tag{6.1}$$

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

$$\partial y_t / \partial x_i' |_{\bar{x}, \bar{y}} = -(\partial F / \partial x_i) / (\partial F / \partial y_t), \tag{6.2}$$

where the right-side partial derivatives are also evaluated at $(\bar{x}, \bar{y})$. Suppose we run the Charnes–Cooper test with $(\bar{x}, \bar{y})$ as the point being tested. Then the optimal dual variables corresponding to the input $\bar{x}_i$ and output $\bar{y}_t$ are respectively $(\partial F / \partial x_i)|_{\bar{x}, \bar{y}}$ and $(\partial F / \partial y_t)|_{\bar{x}, \bar{y}}$. Thus, the rate of change of output y_t 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_t constraint variable![11]

[11] See also Charnes, Cooper and Rhodes (1978, p. 439) for a discussion which can *now* relate this development to the ordinary conditions of economic theory for equality between ratios of marginal productivities and marginal rates of substitution.

More specifically, all Pareto-efficient (X_j, Y_j) of the facet for the point $(\bar{x}, \bar{y})$ satisfy

$$\mu^{*T}y - \nu^{*T}x - \varphi^* = 0, \tag{6.3}$$

where (μ^{*T}, ν^{*T}) are the dual evaluators at an optimal basic solution, *since they do not depend on the Charnes–Cooper test right-hand sides*. Thereby,

$$F(x, y) = \mu^{*T}y - \nu^{*T}x - \varphi^* = 0. \tag{6.4}$$

Clearly, $\mu_t^* = \partial F/\partial y_t$ and $-\nu_i^* = \partial F/\partial x_i$ as already stated.

It should be borne in mind, of course, that these rates of change are valid only for changes which keep one within the facet.

6.2. Isotonicity and economies of scale

To date the structure, or 'geometry', of *empirical* Pareto-efficient production functions has received little attention. The structure depends on and varies with the 'production possibility' or 'reference' set chosen. Here we make a beginning for our new set Q_E and leave to later research more in-depth and broader explorations including those for other reference sets.

In many practical situations we try to choose inputs and outputs with the thought that the underlying empirical Pareto-efficient function should be 'isotone' (which means 'order-preserving'). By definition a (vector) function $f(x)$ is isotone if $x^a \leq x^b$ implies $(f(x^a) \leq f(x^b)$.[12] What in fact can be the case? We show here that in the single-output situation the empirical Pareto-efficient function is *always* isotone. The multiple-output situation, however, may only satisfy a weaker function property which we shall call 'c-d-isotone' or 'cone-directional-isotone', i.e., there is a cone of directions in output space on which the outputs projection is isotone.

Consider first the 'facets' of an empirical Pareto-efficient function for our reference set. These consist of convex combinations of the Pareto-efficient sample input–output points with respect to this reference set.

Theorem 3. Every facet of the empirical Pareto-efficient function relation consists of Pareto-efficient input–output points.

Proof. Each facet corresponds to a basic optimal solution to the following linear program (Charnes–Cooper test) for some sample input–output point

[12] The mathematical term 'isotone' is synonymous with the expression 'monotonically increasing'.

(X_0, Y_0):

$$\min_{\lambda, s^+, s^-} \; -e^{\mathrm{T}}s^+ - e^{\mathrm{T}}s^-,$$

subject to

$$Y\lambda - s^+ = Y_0, \quad -X\lambda - s^- = -X_0, \quad e^+\lambda = 1,$$

$$\lambda, s^+, s^- \geq 0,$$

where $Y \triangleq (Y_1, \ldots, Y_n)$, $X \triangleq (X_1, \ldots, X_n)$, with optimal $s^{*+} = 0$, $s^{*-} = 0$. Every (X_j, Y_j) in the optimal basis is Pareto-efficient since (1) the optimal dual evaluators determined by the basic solution do not depend on the right-hand-side vector $(-X_0, Y_0)$, and (2), replacing $(-X_0, Y_0)$ by any $(-X_j, Y_j)$ in the basis preserves the feasibility of the basis for solution with the new right-hand side.

Next, if we Charnes–Cooper test any convex combination of the basic $(-X_j, Y_j)$, e.g., $(-X_B, Y_B)^{\mathrm{T}}\theta_B$ with $\theta_B \geq 0$, $e_B^{\mathrm{T}}\theta_B = 1$ for Pareto-efficiency and this reference set, i.e., insert it in place of $(-X_0, Y_0)^{\mathrm{T}}$, an optimal λ^* is then simply θ_B (plus $s^{*+}, s^{*-} = 0$) with this same basis. Thus the whole facet is Pareto-efficient. ■

Now we prove:

Theorem 4. If a Pareto-efficient empirical production function has only a single output, then it is an isotone function.

Proof. Suppose $x^a \leq x^b$ for Pareto-efficient (x^a, y^a) and (x^b, y^b). By definition of Pareto-efficiency for (x^b, y^b), some component output y_j^a can exceed y_j^b only if some other output $y_k^a < y_k^b$. Since there is only one output, then $y^a \triangleq f(x^a) \leq y^b \triangleq f(x^b)$. Thus $x^a \leq x^b$ implies $f(x^a) \leq f(x^b)$ for Pareto-efficient points, i.e., the isotone property. ■

To show that a multiple-output Pareto-efficient production function need not be isotone, consider the following one-input, two-output example with three sample points a, b, c:

	a	b	c
x_1	1	2	3
y_1	5	2	1
y_2	5	7	4

Input–output points a and b are Pareto-efficient, c is not. We do *not* have isotonicity since $x^a < x^b$, but $f_1(x^a) \triangleq y_1^a > f_1(x^b) \triangleq y_1^b$.

If now we 'project' the outputs of a and b along the direction given by w_1 and w_2 for y_1 and y_2 we obtain the *single* output $5w_1 + 5w_2$ for a and $2w_1 + 7w_2$ for b. Requiring

$$7w_2 + 2w_1 \geq 5w_2 + 5w_1$$

implies $w_2 \geq \frac{3}{2}w_1$ as the cone of directions (w_1, w_2) which yields an isotonic relation.

In general when $x^c \geq x^d$, if $y_r^c \geq y_r^d$ for $r \in R^+$, where $R^+ \cup R^- = \{1, 2, \ldots, s\}$, then the cone of isotonic directions $(w_1, \ldots, w_s)$ is specified by

$$\sum_{r \in R^+} (y_r^c - y_r^d) \geq \sum_{r \in R} w_r (y_r^d - y_r^c). \tag{6.5}$$

Homogeneous production functions play an important role in the economics literature. 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 'returns 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 $\rho\bar{x}$ in a small neighborhood of $\bar{x}$. Banker, Charnes and Cooper (1984) give a criterion for deciding this (with production possibility set $\mathscr{Q}_E \cup \mathscr{B} \cup \mathscr{C}$ or $\mathscr{Q}_E \cup \mathscr{B}$), but does not give us the rates of change.

Because of our preceding theorems, however, we know that empirical Pareto-efficient functions are c-d-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 $e^{\mathrm{T}}x$. 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 fig. 2 an isotonic function is plotted together with the resulting concave cap (large dashed lines) obtained as the empirical function.

As suggested in our original paper,[13] non-concavity can be explored by applying (output) component by component strictly concave transformations g_r to obtain $g_r(y_r)$ instead of y_r so that $g_r(y_r(x))$ would be concave and out plot might look like fig. 3.

[13]'An Empirical DEA Production Function' by Charnes, Cooper and Seiford, April 1981, CCS 396, Center for Cybernetic Studies, The University of Texas, Austin, TX.

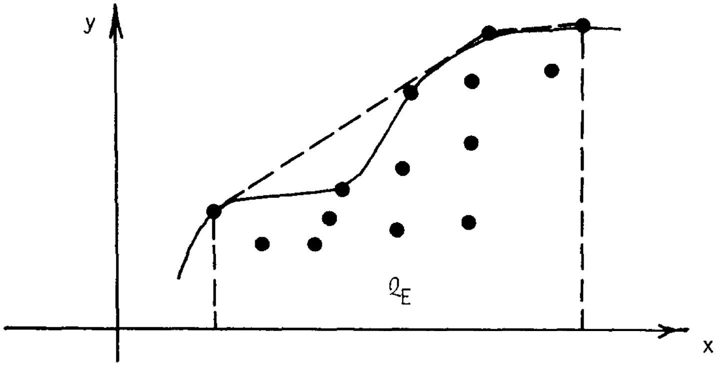

Fig. 2

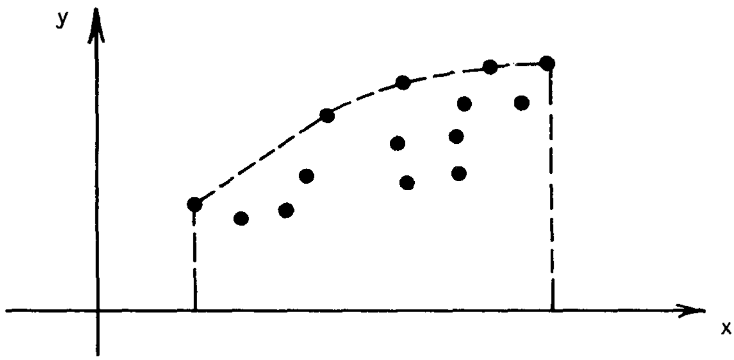

Fig. 3

Informatically, we can do this by applying transformations of form $g_t(y_t) = \bar{y}_t + (y_t - \bar{y}_t)^{1/\beta}$ with $\beta \approx 20$ to obtain possible new facets in the $g_t(y_t)$ (where $\bar{y}_t = \min_j y_{tj}$).

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.

6.3. 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.[14] 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 (4.5) 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.

7. 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-efficiency characterizations and constructions of empirical production functions restrict us methodologically to exploration of such functions by means of concave 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-efficient constructions should also be clear.

References

Afriat, S., 1972, Efficiency estimation of production functions, International Economic Review 13, 568–598.

Aigner, D.J., C.A.K. Lovell and P.J. Schmidt, 1977, Formulation and estimation of stochastic frontier production function models, Journal of Econometrics 6, 21–37.

Banker, R., 1984, Estimating most productive scale size using data envelopment analysis, European Journal of Operational Research 17, 35–44.

Banker, R., A. Charnes and W.W. Cooper, 1984, Models for estimating technical and scale efficiencies in data envelopment analysis, Management Science 30, 1078–1092.

[14] See Charnes and Cooper (forthcoming).

Banker, R., W. Bowlin, A. Charnes and W.W. Cooper, in process, Mix inefficiencies and synergies in data envelopment analysis, CCS report (Center for Cybernetic Studies, The University of Texas, Austin, TX).

Ben-Israel, A., A. Ben-Tal and A. Charnes, 1977, Necessary and sufficient conditions for a Pareto optimum in convex programming, Econometrica 45, 811–820.

Bessent, A., W. Bessent, A. Charnes, W.W. Cooper and N. Thorogood, 1983, Evaluation of educational program proposals by means of DEA, Educational Administration Quarterly 19, 82–107.

Charnes, A. and W.W. Cooper, 1962, Programming with linear fractional functionals, Naval Research Logistics Quarterly 9, nos. 3, 4.

Charnes, A. and W.W. Cooper, 1977, Goal programming and multiple objective optimizations, European Journal of Operational Research 1, no. 1.

Charnes, A. and W.W. Cooper, 1961, Management models and industrial applications of linear programming (Wiley, New York).

Charnes, A. and W.W. Cooper, 1980, Auditing and accounting for program efficiency in not-for-profit entities, Accounting, Organizations and Society 5, 87–107.

Charnes, A. and W.W. Cooper, forthcoming, Preface to topics in data envelopment analysis, Annals of Operations Research.

Charnes, A., W.W. Cooper and E. Rhodes, 1978, Measuring the efficiency of decision-making units, European Journal of Operations Research 2, 429–444.

Charnes, A., W.W. Cooper and E. Rhodes, 1981, Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through, Management Science 27, 668–687.

Charnes, A., W.W. Cooper and A. Schinnar, 1982, Transforms and approximations in cost and production function relations, Omega 10, 207–211.

Charnes, A., W.W. Cooper and D. Sherman, in process, A comparative study of DEA with ratio and regression approaches to efficiency evaluations, CCS report (Center for Cybernetic Studies, The University of Texas, Austin, TX).

Charnes, A., W.W. Cooper, L. Seiford and J. Stutz, 1982, A multiplicative model for efficiency analysis, Socio-Economic Planning Sciences 6, 223–224.

Charnes, A., W.W. Cooper, L. Seiford and J. Stutz, 1983, A units invariant multiplicative measure, Operations Research Letters.

Charnes, A., W.W. Cooper, A. Lewin, R. Morey and J. Rousseau, forthcoming, An approach to positivity and stability analysis in DEA, Annals of Operations Research.

Färe, R. and C.A.K. Lovell, 1978, Measuring the technical efficiency of production, Journal of Economic Theory 19, 150–162.

Farrell, M.J., 1957, The measurement of productive efficiency, Journal of Royal Statistical Society A 120, 253–281.

Fenchel, W., 1953, Convex sets, cones and functions (Princeton University Press, Princeton, NJ).

Försund, F. and L. Hjalmarsson, 1979, Generalized Farrell measures of efficiency: An application to milk processing in Swedish dairy plants, Economic Journal 89, 294–315.

Frisch, R., 1935, The principles of substitution: An example of its application in the chocolate industry, Nordisk Tidsskrift for Teknisk Ekonomi 1, 12–27.

Rockfellar, T., 1970, Convex analysis (Princeton University Press, Princeton, NJ).

Shephard, R.W., 1953, Cost and production functions (Princeton University Press, Princeton, NJ).

Shephard, R.W., 1970, Theory of cost and production functions (Princeton University Press, Princeton, NJ).

Schaible, S, 1974, Parameter-free convex equivalent and dual programs of fractional programming problems, Zeitschift für Operations Research 18, 187–196.