

AD-A038 211

TEXAS UNIV AT AUSTIN CENTER FOR CYBERNETIC STUDIES
MEASURING THE EFFICIENCY OF DECISION MAKING UNITS WITH SOME NEW--ETC(U)
OCT 76 A CHARNES, W W COOPER, E RHODES
N00014-75-C-0616
NL

UNCLASSIFIED

1 OF 1
AD
A038211

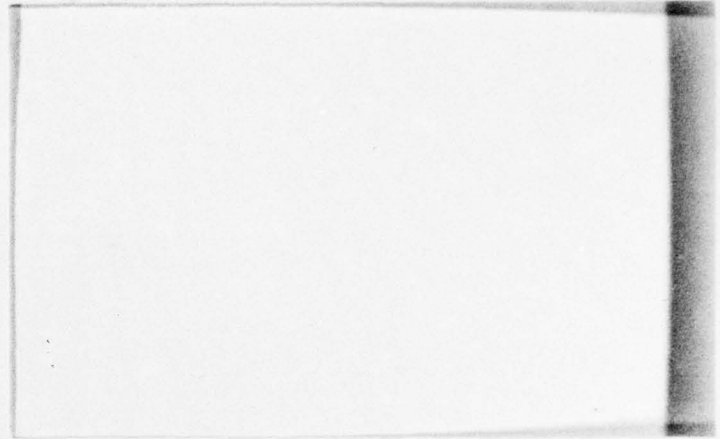


END

DATE
FILMED
4-77

ADA 038211

12
b5



1473

CENTER FOR CYBERNETIC STUDIES

The University of Texas
Austin, Texas 78712

DDC
RECEIVED
APR 19 1977

Q7

B



DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

Research Report CCS 276

MEASURING THE EFFICIENCY OF
DECISION MAKING UNITS WITH SOME NEW
PRODUCTION FUNCTIONS AND
ESTIMATION METHODS

by

A. Charnes*
W.W. Cooper**
E. Rhodes***

ACCESSION NO.	
NTIS	White Section ☒
DDC	Diff Section ☐
UNANNOUNCED	☐
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	AVAIL. 000/00 SPECIAL
A	

October 8, 1976

*The University of Texas.

**Graduate School of Business Administration, Harvard University.

***School of Urban and Public Affairs, Carnegie-Mellon University.

This paper represents a revision of an earlier version entitled "Exposition, Interpretation, and Extensions of the Farrell Efficiency Measure".

This research was partly supported by Project NR047-021, ONR Contract N00014-75-C-0616 with the Center for Cybernetic Studies, The University of Texas, and ONR Contract N00014-76-C-0932 at Carnegie-Mellon University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

CENTER FOR CYBERNETIC STUDIES

A. Charnes, Director
Business-Economics Building, 203E
The University of Texas
Austin, Texas 78712
(512) 471-1821

406177

LB

Abstract

A new series of linear programming models is used to clarify and extend a measure of efficiency introduced by M. J. Farrell. Multiple output as well as single output versions are provided which relate this efficiency to ordinary Pareto-Koopmans optimality conditions. Their duals supply numerical estimates of production coefficients. This duality, from mathematical programming, is distinguished from another in the theory of production (as found in the works of Samuelson, Shephard, et al) and then used to provide new nonlinear duals which open contacts with other mathematical programming developments in areas like fractional programming, etc.

Introduction

In a pioneering study [1] , M. J. Farrell provided a new approach to the measurement of productive efficiency from empirical data. In that study, and in subsequent ones as well,^{1/} he was concerned to provide a measure which reflected all of the pertinent factors utilized by different decision making units in producing their outputs. Thus, in order to distinguish Farrell's measures from the more customary (one at a time) measures of factor productivity -- e.g., measure of labor or capital productivity -- we will refer to these Farrell-type measures as measures of efficiency for "decision making units." These units will hereafter be referred to as "firms." The latter term is to be accorded only generic significance however since it is intended to cover public service (not for profit) organizations as well as internal subdivisions or even activities which might not ordinarily be regarded as separately identified entities. The point at issue in any case is only whether they make decisions that involve one or more resource combinations and, if so, whether these decisions are made efficiently or not.

We might note that Farrell's discussion and examples point toward situations like international trade or regulated industries where ordinary market-price considerations have only limited value. That is, here, as elsewhere in his work, Farrell used economic principles and theory as his main guide. He also turned attention to the need for methods of estimating "efficiency frontiers" from empirical data. In the course of

^{1/} See [19].

doing so he turned to a variety of mathematical techniques to secure his estimates and thereby replaced a previous (almost) exclusive emphasis on statistical "averaging" methods such as least squares, etc., in order to obtain these frontier estimates to which his economic constructs accorded the primacy. At the same time he also kept the latter, i.e., the economics constructs, in perspective by focussing attention on "relative efficiency", as measured from actually observed behavior instead of trying to deduce his measures from a best possible efficiency via purely theoretical considerations.

As might be expected, at the early stages of such developments, there are a variety of ambiguities in Farrell's treatment that still require attention. We shall try to supply this needed clarification. In the course of doing so, we shall also eliminate the need for a variety of constructs like the "points of infinity" which Farrell (and others following him) employed despite their awkward character -- both in mathematics and economics.

This will all be done by a suitably erected series of linear programming problems. These will be used first for the indicated clarifications in the single output case. These linear programming models will then be extended to the case of multiple outputs in order thereby to provide a uniform approach to this and the preceding (single output) case.

We should therefore note that Dr. Alan Hoffman has also suggested a linear programming approach and that he did so at the original session in which Farrell's paper was first presented.^{1/} Hoffman's formulation, which was restricted to the single output case, however, differs from ours in that (a) it is presented in equality form, i.e., he does not deduce our inequality form, and (b) it carries forward with it such

^{1/} See pp. 284-285 in [18].

concepts as "points at infinity" by approximation with "large" real multiples of unit vectors. Hoffman's suggested approach was also, apparently, differently motivated in that it was directed only to issues of computation^{1/} and did not attend to the kinds of theoretical inquiries and extensions to coefficient estimation techniques which are of concern in the present paper. Our models, on the other hand, are formulated to provide access to a related series of dual problems which, as we shall see, provide a variety of new openings for empirical studies of production behavior.

Inter alia, these models thereby provide access to new methods for securing empirical estimates of isoquant coefficients (as represented in piecewise linear facets) from empirical data for the single output case as well as the activity vector coefficients and related production possibility surfaces in multiple output cases. Furthermore, together with their related primal problems they also provide access to a variety of new production function possibilities that might be erected for different kinds of inquiries. At one level such a function might be regarded as an individual firm production function of a "representative efficient firm" variety.^{2/} At another level they lead to a variety of

^{1/} In an effort to eliminate the laborious and costly matrix inversion routines which Farrell had been employing, Dr. Hoffman suggested a use of the then new "dual method" of C. E. Lemke, as we shall illustrate below, and, in fact, Farrell and others subsequently used these procedures and found them very effective for single output investigations. See, e.g., the discussion on pp. 264-267 of [19].

^{2/} Insofar as the English language can bear such burdens, we should probably utilize the plural rather than the singular for this "representative efficient firm" concept in order thereby to emphasize also that these efficient facets alter in "slope", e.g. in normal vector, in piecewise constant fashion.

additional possibilities depending on the types of aggregation/dis-aggregation relations which might be wanted for studying the underlying decision making units, and so on. Finally, they also provide access to a series of nonlinear fractional programming problems which, in turn, provide access to still further relations of duality for future research as well as immediate interpretations for the related programming problems and production function possibilities covered in the present paper.

We shall try to develop these ideas first in a fairly leisurely and informal manner. This will be done numerically and verbally for the simple single output case. Relations with other uses of dual (we shall call them complementary) problem approaches in the theory of production will then be briefly sketched. With this for background we shall next proceed to more straightforward mathematical developments for the subsequent generalizations that we shall make in a more rapid coverage. We shall then conclude the body of the article by discussing some of the further avenues of research (and application) that these developments appear to offer. An appendix is also provided which deals with issues such as the invariance of Farrell's measure of efficiency.

Background Illustration

At this point we should now observe that Farrell actually introduced a variety of measures^{1/} from which we (like Farrell) will focus on only one, which he called "technical efficiency". For this measure it is assumed only that all inputs and outputs have value. They are not specifically priced or costed so that this measure of efficiency, as we shall see, actually rests on the concepts of Pareto-Koopmans^{2/} optimality in which output increases (or input

^{1/} See Farrell [18] and see also [20] for further discussion.

^{2/} See Chapter IX in [9].

decreases) are obtainable only if they do not reduce other outputs or increase other inputs.

Thus, bearing in mind that we are here concerned with technical efficiency only -- and not the related concepts of price, profit, or even "overall" efficiency^{1/} -- we turn to Figure 1 for our simple hypothetical case. To provide our illustrative initial example, we now proceed as follows: First we reserve P_6 for separate discussion. Then we interpret $P_1, \dots, P_5$ as points in a two-dimensional space in which the coordinates represent observed amounts of two different factors of production used to produce a single product by five different firms.

We are here concerned with the one output, as distinguished from the multiple output, case. We do have multiple inputs (here two), however, and, given these observations, the first problem is to locate the efficiency frontiers or rather the subset of extreme points from which such a frontier may be formed by reference to the five points, P_1, P_2, P_3, P_4 , and P_5 as shown. These "efficient" points are then joined by linear segments, as in Figure 1, to form the corresponding "efficient isoquant" which, as can be seen, is a "piecewise linear" and continuous curve. Here in this two-dimensional space the efficient surface is a line, i.e., an isoquant line, but in m -dimensions, "facets" of dimension greater than one will be subtended in an efficient isoquant surface.

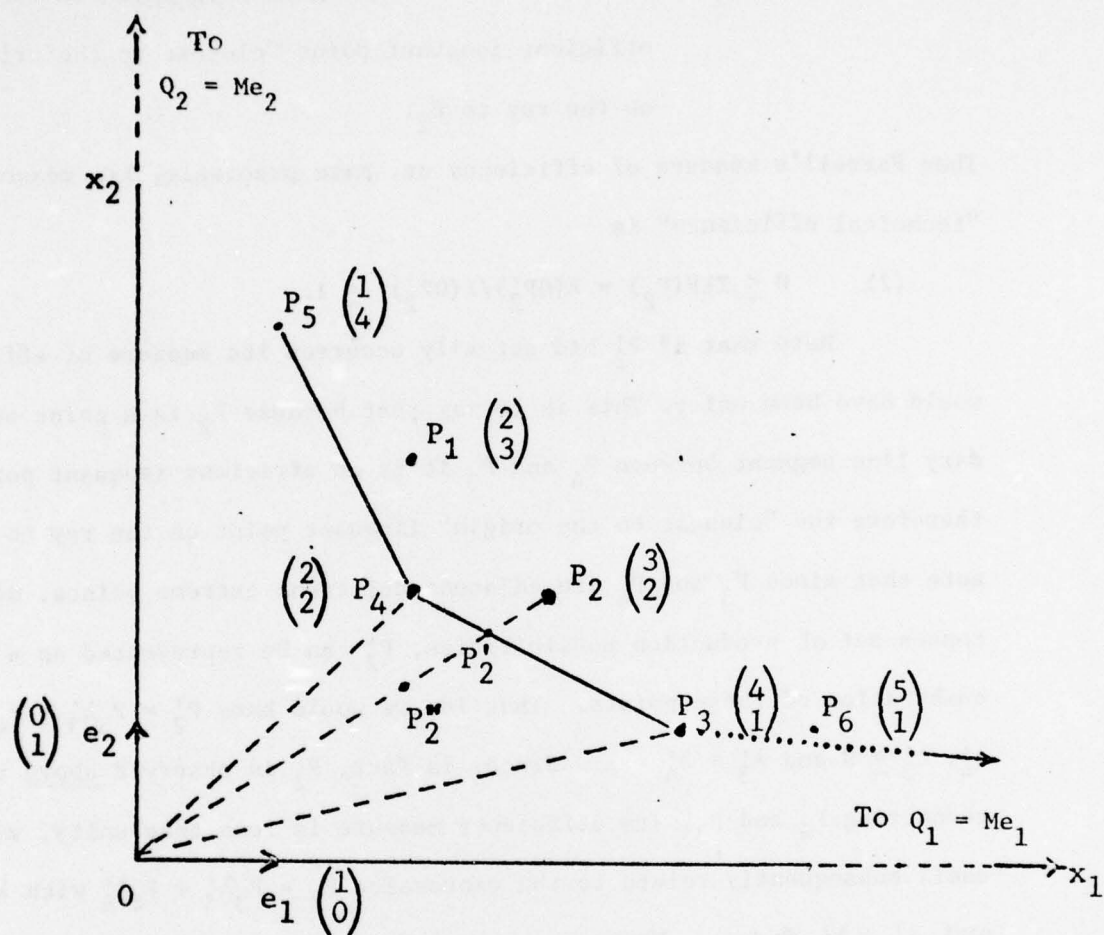
^{1/} See Farrell [18] for his discussion of "price efficiency", "overall efficiency", etc. See also Forsund [20]. We should perhaps also note that Farrell himself does not really try to exploit these concepts and in fact emphasizes his skepticism (as do we) on their manageability, except, perhaps, in special circumstances.

An "efficient isoquant" in this analysis represents what Farrell regards as "practically attainable efficiency." By this he means that at least some "firms" attained this level of efficiency even though others did not. Then, by interpolation between these observed points, he provides a "facet" (here a line) of reference for measuring the relative efficiency of the remaining firms.

To develop his measure of efficiency, Farrell then makes two key assumptions. The first is that any firm can expand or contract along a "ray" through the origin such as the one indicated by the broken line from the origin to P_2 . The other assumption is that all convex combinations of the observed points, $P_1, \dots, P_5$ represent actual production possibilities. The "efficient points" then correspond to the extreme points of this convex set plus the union of the boundary facets of this convex set, i.e., the union of the boundary simplexes formed on the efficient extreme points. The latter thus constitutes the "efficient isoquant" or the "efficient facet" - viz., the locus of points corresponding to the minimum inputs of all factors for a specified level (e.g., one unit) of the observed output.

To see what is involved graphically in Farrell's proposed measure refer to Figure 1 where the "efficient isoquant" is drawn in piecewise linear segments from P_5 to P_4 and P_4 to P_3 . These points constitute the input vectors, per unit output, for firms 3, 4, and 5 with linear interpolation specifying what should be attainable for P_1 and P_2 . For example, let

Figure 1



$\ell(OP_2) = \text{length of "ray from origin" to } P_2$

(1) and

$\ell(OP'_2) = \text{length of ray from origin to } P'_2 \text{ which is the}$
 efficient isoquant point "closest to the origin"
 on the ray to P_2 .

Then Farrell's measure of efficiency or, more precisely, his measure of "technical efficiency" is

$$(2) \quad 0 \leq \text{TEF}(P_2) = \ell(OP'_2) / \ell(OP_2) \leq 1.$$

Note that if P'_2 had actually occurred its measure of efficiency would have been unity. This is to say that because P'_2 is a point on the boundary line segment between P_4 and P_3 it is an efficient isoquant point and therefore the "closest to the origin" isoquant point on the ray to P'_2 . We note that since P_3 and P_4 are adjacent efficient extreme points, of the convex set of production possibilities, P'_2 can be represented as a convex combination of these points. That is, we would have $P'_2 = P_3\lambda'_3 + P_4\lambda'_4$ with $\lambda'_4, \lambda'_3 \geq 0$ and $\lambda'_3 + \lambda'_4 = 1$. Since, in fact, P_2 is observed above the line connecting P_4 and P_3 , its efficiency measure is less than unity, which we shall subsequently relate to the expression $P_2 = P_3\lambda'_3 + P_4\lambda'_4$ with $\lambda'_3, \lambda'_4 \geq 0$ and $\lambda'_3 + \lambda'_4 \geq 1$. Moving next to P_2'' we can also express this, as $P_2'' = P_3\lambda'_3 + P_4\lambda'_4$, with $\lambda'_3, \lambda'_4 \geq 0$, $\lambda'_3 + \lambda'_4 < 1$. In each case non-negativity of both variables is preserved with the resulting sum indicating where the point is relative to the isoquant segment connecting P_3 and P_4 .

Now consider non-negativity. All points inside the cone with edges indicated by the broken line from the origin through P_3 and P_4 , respectively, can be expressed as nonnegative combinations of these two points. Only when one attempts to express a point outside the cone, e.g., a point such as P_1 , is the nonnegativity condition violated. But then a recourse to our piecewise linearity assumption

allows us to generate a new isoquant segment connecting P_4 and P_5 from which to evaluate the relative efficiency of points such as P_1 , and so on.

With this synopsis we can initiate our discussion of the suggested measure by indicating its relation to "actual" resource utilization. The idea is to provide a measure which will show the amount of resource reduction which could be effected for a point such as P_2 if it could be moved to a point on the "efficient isoquant". In fact, the measure provided in (2) shows the proportionate reduction in both x_1 and x_2 by moving along the ray from the origin to intersection with the isoquant connecting P_3 and P_4 . Thus, assuming that such a convex combination of P_3 and P_4 were available, then this hypothetical firm could produce one unit of output with less of each resource than was observed for P_2 .

So far, we have focused on economics and the related technological concepts and considerations. What about the statistical concepts that also enter into the consideration of any observations from which inferences are to be drawn? Farrell boldly, but we think knowingly, flies into the face of a long tradition and insists that it is economic theory rather than statistics which should be given primacy and, following his lead, we will do the same while allowing for any help that statistics may provide in dealing with these essentially economic considerations. Via deterministic mathematical models and methods we can then in economics, as in other disciplines, perhaps proceed to develop a variety of additional approaches which may be of value for important empirical applications with particular reference to potential uses of these economic constructs for a variety of social policy problems. In this way we can perhaps provide a broader basis for empirical economics and thereby augment its potential value for various important applications.

Bearing this all in mind, we now formalize from the above illustration as follows: For each $j = 1, \dots, n$ firms we are given observations on their inputs for each of two factors of production which we symbolize as

$$(3.1) \quad \begin{aligned} x_{1j} &= \text{amount of first factor used by firm } j \\ x_{2j} &= \text{amount of second factor used by firm } j \end{aligned}$$

We are also given the amount of output -- sales, shipments, or production -- which we symbolize as

$$(3.2) \quad y_j = \text{amount of output of firm } j.$$

In each case we assume that the amounts in (3.1) and (3.2) are all positive and then combine them for ratio representation, as in Figure 1, via

$$(4) \quad \begin{aligned} x'_{1j} &= x_{1j}/y_j \\ x'_{2j} &= x_{2j}/y_j \end{aligned}$$

In other words, x'_{1j} , x'_{2j} are, respectively, the inputs per unit output utilized by firm $j = 1, \dots, n$. The points P_j in Figure 1 are then represented as

$$(5) \quad P_j = \begin{bmatrix} x'_{1j} \\ x'_{2j} \end{bmatrix}$$

Thus the observations in Figure 1 are all "normed" by reference to the outputs for, respectively, each of the $j = 1, \dots, 5$ firms we are presently considering. Accordingly, the "efficient isoquant" exhibited there represents the required input rates for 1 unit of output and so we also refer to it as the "unit isoquant."

We can now conclude this section by further formalizing some of our preceding discussion as follows. Consider any two points P_r and P_s which form a basis for the 2-dimensional space exhibited in Figure 1. That is we assume P_r and P_s , which are 2-component vectors, are linearly independent. Any other observed point, P_o , say, may be expressed in terms of them via

$$(6.1) \quad P_r \lambda'_r + P_s \lambda'_s = P_o.$$

Now assume P_r and P_s are also efficient. If $\lambda'_r, \lambda'_s \geq 0$ then P_o is in the cone with edges formed by extending a ray from the origin through each of P_r and P_s , respectively. Conversely, if P_o is not in that cone then at least one of the

variables required for the expression of P_o in (6.1) must be negative. If

$$(6.2) \quad \lambda'_r, \lambda'_s \geq 0$$

and

$$(6.3) \quad \lambda'_r + \lambda'_s = 1$$

then P_o is on the line segment connecting P_r and P_s . If, on the other hand, $\lambda'_r + \lambda'_s > 1$ then P_o cannot be efficient while if $\lambda'_r + \lambda'_s < 1$ then P_r and P_s cannot both be efficient.

In any case since we are considering points on a ray through the origin, their coordinates in terms of any basis are proportional. Thus we have

$$(6.4) \quad 0 \leq \text{TEF}(P_o) = 1/(\lambda'_r + \lambda'_s) \leq 1$$

with (6.2) and (6.3) also applying whenever P_r , P_s and P_o are all efficient and in the cone.

Linear Programming Formulations

Now consider the case where there are n firms and m inputs so that

$$(7.1) \quad P_j = \begin{bmatrix} x'_{1j} \\ \vdots \\ x'_{ij} \\ \vdots \\ x'_{mj} \end{bmatrix}$$

where

$$(7.2) \quad x'_{ij} = x_{ij}/y_j$$

$$i = 1, \dots, m$$

$$j = 1, \dots, n,$$

which means that each of these input amounts is normed on y_j , the observed output for the j th firm, as before.

P_o is assumed to be a vector of observations associated with one of $j = 1, \dots, n$ firms whose inputs, per unit output, have been observed. For concreteness let $P_o = P_k$ and let it be desired to determine whether P_k is efficient. To check on this we can form the linear programming problem,

$$(7.3) \quad \max z_o = \sum_{j=1}^n \lambda'_j$$

$$\text{with} \quad \sum_{j=1}^n P_j \lambda'_j = P_o, \lambda'_j \geq 0$$

and, to insure that we consider P_o relative to a cone on efficient extreme points which contains P_o , we employ an adjacent extreme point algorithm such as the "simplex" or "dual" method to solve this linear programming problem.

We then have as one possible solution

$$(8) \quad P_k \lambda'_k = P_k$$

with $\lambda'_k = 1$. It follows that $\max z_o = z_o^* \geq 1$ and therefore

$$(9) \quad 0 \leq \text{TEF}(P_o) = 1/(\sum_{j=1}^n \lambda'^*_j) = 1/z_o^* \leq 1$$

as required.

We need to insure that the efficient isoquant, when attained, accommodates all possibilities. For instance, we need to insure that some firm using an extremely large amount of one factor and a very small amount of every other factor will still be on a ray which intersects the efficient isoquant en route to the origin. Farrell adds "points at infinity" to insure this and, following him, Hoffman also does the same. Now, however, we show how all such possibilities can be covered in a new formulation free of such constructs.

We can represent the points at infinity as $Q_i = M e_i$, $i = 1, \dots, m$ where M is a non-Archimedean transfinite number ($M > r$, r any real number)^{1/} and the e_i are the unit coordinate vectors - see, e.g., e_1 and e_2 in Figure 1. The other points P_j have finite and non-negative entries. We now consider the non-negative representation of P_0 , a finite, non-negative vector, in terms of the P_j and Q_i - e.g.,

$$\begin{aligned} P_0 &= \sum_{j=1}^n P_j \lambda'_j + \sum_{i=1}^m Q_i \xi_i \\ (10.1) \quad &= \sum_{j=1}^n P_j \lambda'_j + M \sum_{i=1}^m e_i \xi_i \end{aligned}$$

where

$$\lambda'_j, \xi_i \geq 0.$$

Since the entries in the P_j and e_i are finite and non-negative, such a representation is possible in terms of finite λ'_j (if and) only if the ξ_i are of the form

$$(10.2) \quad \xi_i = \frac{1}{M} s_i$$

where the s_i are finite and non-negative. Thus, we get the equivalent finite representation of the constraints as

$$\begin{aligned} (10.3) \quad &\sum_{j=1}^n P_j \lambda'_j + \sum_{i=1}^m e_i s_i = P_0 \\ &\lambda'_j, s_i \geq 0. \end{aligned}$$

To complete this part of the analysis we now explicitly formulate the linear programming problem with points at infinity as follows,

^{1/} See, e.g., Robinson [29].

$$\begin{aligned}
 \text{Max } z_o &= \sum_{j=1}^n \lambda'_j + \sum_{i=1}^m \xi_i \\
 (11.1) \quad \text{with } P_o &= \sum_{j=1}^n P_j \lambda'_j + \sum_{i=1}^m Q_i \xi_i \\
 \lambda'_j, \xi_i &\geq 0.
 \end{aligned}$$

Then we replace this with the following new form

$$\begin{aligned}
 \text{Max } z_o &= \sum_{j=1}^n \lambda'_j + \frac{1}{M} \sum_{i=1}^m s_i \\
 (11.2) \quad \text{with } P_o &= \sum_{j=1}^n P_j \lambda'_j + \sum_{i=1}^m e_i s_i \\
 \lambda'_j, s_i &\geq 0,
 \end{aligned}$$

where we also note that the constraints can be written in our newly established equivalent inequality form as

$$\begin{aligned}
 (12) \quad P_o &\geq \sum_{j=1}^n P_j \lambda'_j \\
 \lambda'_j &\geq 0, j = 1, \dots, n.
 \end{aligned}$$

This inequality format for the constraints can be given a geometric interpretation in terms of relative efficiency of two production possibilities - namely, P_i is at least as efficient as P_j if $P_i \leq P_j$. In this manner the Farrell efficient (or frontier) points are simply "vector" or "Pareto minimal" points of the set of all production possibilities. This means that P_j is not efficient relative to P_i if P_i contains at least one entry strictly less than the corresponding entry in P_j .

A further geometric interpretation of the set of production possibilities can also be made. It is a convex set derived from the convex hull of the observed points by adding to this convex hull all points northeast of it. Thus, if P_r is a production possibility then so is P_s if $P_s \geq P_r$.

The non-Archimedean programming problem (11.2) provides us with a z_o^* whose reciprocal is the Farrell efficiency measure up to a (non-Archimedean) infinitesimal. If any of the s_i - which may be considered as slack variables for the representation (12) - is in at positive value in an optimal solution, then P_o cannot be efficient. If all s_i are at zero value, then the infinitesimal part of z_o^* is zero. Thus, we can make the following alternate characterization of efficiency,

$P_k = P_o$ is efficient if and only if

- (13) i. Its optimal z_o value is $z_o^* = 1$, and
 ii. The slack variables are all at zero value in every optimum tableau,

for the problem

$$\begin{aligned} \text{Max } z_o &= \sum_{j=1}^n \lambda'_j \\ \text{with} \\ (14.1) \quad \sum_j P_j \lambda'_j &\leq P_o \\ \lambda'_j &\geq 0 \end{aligned}$$

or its equivalent

$$\begin{aligned} \text{Max } z_o &= \sum_{j=1}^n \lambda'_j \\ \text{with} \\ (14.2) \quad \sum_{j=1}^n P_j \lambda'_j + \sum_{i=1}^m e_i s_i &= P_o \\ \lambda'_j, s_i &\geq 0 \end{aligned}$$

Computational Example

We illustrate the above developments via the data of Figure 1. To initiate this illustration we shall first consider only the points P_1 , ..., P_5 and reserve P_6 for separate consideration. This leads to the tabular arrangement in Table 1 where, with the conventions of [9], we use B and c_B to designate basis vectors and their functional coefficients at each stage. Thus, at "Stage 0" the basis B consists of the slack vectors e_1 and e_2 with functional coefficients of zero shown explicitly on the left.

(Insert Table 1)

Using full (rather than contracted) tableau arrangements for effecting our simplex calculations, we proceed from "Stage 0" to an optimum at "Stage 3*" from which we secure the following information: P_4 and P_5 are in the optimal basis and hence are both efficient. Concomitantly, $P_0 = P_1$ has $z_0^* = 7/6 > 1$ and hence is not efficient - see (13) - and, in fact,

$$TEF(P_0 = P_1) = 1/z_0^* = 6/7$$

measures the reduction in factor inputs which could have put this firm on the unit isoquant if it had been producing efficiently - viz.,

TABLE 1

Illustration of Simplex and Dual Method Calculations

SIMPLEX METHOD

Stage	Basis		Structural Vectors					Slack Vectors		Stipulations Vector	Optimal Basis
	c_B	B	P_1	P_2	P_3	P_4	P_5	e_1	e_2	$P_0 = P_1$	
0	0	e_1	2	3	4	2	1	1		2	
	0	e_2	3	2	1	2	↑ 4		1	3	
	$z_j - c_j$		-1	-1	-1	-1	↓ -1				
1	0	e_1	5/4	5/2	15/4	3/2		1	-1/4	5/4	
	→ 1	P_5	3/4	1/2	↑ 1/4	1/2	1		1/4	3/4	
	$z_j - c_j$		-1/4	-1/2	↓ -3/4	-1/2			1/4	3/4	
2	→ 1	P_3	1/3	2/3	1	2/5		4/15	-1/15	1/3	
	1	P_5	2/3	1/3		↑ 2/5	1	-1/15	4/15	2/3	
	$z_j - c_j$					↓ -1/5		1/5	1/5	1	
3*	→ 1	P_4	5/6	5/3	5/2	1		2/3	-1/6	5/6	P_4, P_5
	1	P_5	2/6	-1/3	-1		1	-1/3	2/6	2/6	
	$z_j - c_j$		1/6	1/3	1/2			1/3	1/6	7/6	

DUAL METHOD

DUAL METHOD

	1	P ₄	5/6	5/3	5/2	1		2/3	-1/6	5/3
← 4	1	P ₅	2/6	-1/3	-1		1	-1/3	2/6	-1/3
	z _j - c _j		1/6	1/3	1/2			1/3	1/6	4/3

P₀ = P₂

	1	P ₄	5/3	5/6	↓	1	5/2	-1/6	2/3	5/6
← 5*	1	P ₃	-1/3	2/6	1		-1	2/6	-1/3	1/3
	z _j - c _j		1/3	1/6		1/2		1/6	1/3	7/6

P₄, P₃

$$6/7P_1 = 6/7 \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 12/7 \\ 18/7 \end{bmatrix} .$$

Since P_4 and P_5 are efficient, there is a possibility that the basis at Stage 3* could also be used to measure the efficiency of other points besides P_1 . In fact, this would be the case if any of the other vectors, P_2 and P_3 , had only non-negative entries in their columns at this tableau stage, i.e., at Tableau Stage 3*. However, this is not the case. Both P_2 and P_3 have negative entries in their columns -- which means that neither they, nor the rays to them (from the origin), lie on or intersect the line segment connecting P_4 and P_5 . In other words, P_3 and P_2 lie outside the cone generated from the origin through P_4 and P_5 and hence cannot be expressed as non-negative combinations of these two points. See Figure 1.

We next observe, with A. J. Hoffman, that we are now in a position to use C. E. Lemke's "dual method".^{1/} This, too, is illustrated in Table 1 by according P_2 the status of a new P_0 and transferring the data for it at Stage 3* to the new tableau represented as "Stage 4". We effect this transfer only for clarity, however, since we could as easily have proceeded directly from Stage 3* without any tableau such as Stage 4 to achieve the new optimum shown at Stage 5*. This is all done via the dual method, as

^{1/}See [27]. For further treatment of the dual method also see Chapter X of [9].

indicated, and hence there is no need to start afresh or even to backtrack to earlier tableaux.

The results in Stage 5* show that P_4 and P_3 are efficient and that P_2 is not efficient. Indeed, P_2 is "exactly" as inefficient as P_1 . Observe, however, that P_1 was referred to P_4 and P_5 while P_2 was referred to P_3 and P_4 , so that the efficient referents are not the same. Also, the term "exactly" must be taken with a "grain of statistical salt" in any case, since there is some room for observational error which would be evident on replication and this, too, must be allowed for, at least qualitatively, when interpreting the results of this kind of deterministic approach^{1/}.

The just illustrated continuation from the simplex to the dual method evidently provides an efficient computational outline for these determinations. It is not likely that improvement can be secured by recourse to other methods such as, e.g., the "double description method" of T. S. Motzkin, et. al. [28]. Hence, we do not elaborate further on Hoffman's original suggestion with respect to any improvement in computational efficiency. We only observe that simple arrangements such as are displayed in Table 1 should provide all that is wanted for the usual case in economics where, ordinarily, only a very small number of constraints will be considered.

^{1/}The reader should also refer to the exchange between M. G. Kendall, M. Quenouille, and M. J. Farrell in [18].

Now we turn to

$$P_6 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

As depicted in Figure 1, this point lies to the right of the efficiency frontier which terminates at P_3 . Via the developments provided in (10.1) ff., we can, of course, exhibit the nonefficiency of P_6 within the tableau for (14.2). We proceed, however, to a separate treatment of P_6 in order (a) to clarify how part ii of (13) enters into the consideration of efficiency and also (b) to show how some of the additional information available in these extended tableaus may be utilized.^{1/}

First we observe that the inverse of any basis is found under the slack vectors at any stage. Thus, continuing with (14.2) at Stage 5* we immediately have

$$B^{-1} = (P_4, P_3)^{-1} = \begin{bmatrix} -1/6 & 2/3 \\ 2/6 & -1/3 \end{bmatrix}$$

from Stage 5* in Table 1. Hence, we also have

$$B^{-1} P_6 = \begin{bmatrix} -1/6 \\ 8/6 \end{bmatrix}$$

for insertion in the " P_0 " column at this stage. As was the case at Stage 4, one of the components is negative, but we are, nevertheless, in a position to continue with the dual method, just as before.^{2/}

^{1/} See [9] and [17] for further discussion.

^{2/} Since P_k , the new vector to be considered, has $P_k \geq P_r$, the usual condition for continuation with the dual method is automatically satisfied. I.e., one has $z_j - c_j \geq 0$, all j , including $j=k$.

One iteration with the dual method produces

$$P_3 \lambda_3^{*'} + e_1 s_1^* = P_0 = P_6$$

with

$$\lambda_3^{*'} = 1 \text{ and } s_1^* = 1$$

and all other variables at zero value. Evidently the zero slack requirement in (13) is violated and so P_6 is not efficient. The interpretation is also relatively straightforward - viz., a reduction of $x_1 = 5$ to $x_1 = 4$ would eliminate this positive slack and bring P_6 into coincidence with P_3 , in which case $\lambda_3^{*'} = z_0^* = 1$ and both conditions in (13) being satisfied efficiency would again be achieved.

There is no ambiguity in the development via (14.2) and ii in (13), as we have just seen. On the other hand, we could have proceeded by (11.2) in which case we would have,

$$z_0^* = 1 + \frac{1}{M} > 1,$$

and so, again, P_6 is not efficient - via the condition " $z_0^* = 1$ if and only if P_0 is efficient," which would be the applicable one for (11.2). Then, as before, we would have

$$0 \leq \text{TEF}(P_6) = \frac{1}{z_0^*} = \frac{1}{1 + \frac{1}{M}} < 1$$

indicating the reduction in both factors required to bring P_6 onto the efficient isoquant segment stretching from P_3 to Q_1 .

Duality and Other Tableau Characterizations

We should stress the importance of employing an adjacent extreme point method in the calculations and characterization of extreme efficient points. In particular, as will be seen, the tableaus resulting from simplex and dual method calculations - which are both adjacent extreme point

methods - are rich in information concerning such things as efficient frontier facets and the normal vectors to them.

In order to map out the efficient isoquant we proceed by determining the extreme points spanning one facet plus the normal to this facet. We then move on through adjacent efficient facets and continue doing so until all of the wanted efficient facets and normal directions have been determined. The latter, as we shall show, correspond to an optimal solution to the dual problem associated with the primal problem formed from (14.1) or (14.2).

Our procedure allows us to characterize these efficient facets explicitly and in ways which have not heretofore been available. Hence, we shall develop them here in all detail and generality as follows.

We shall first assume that all of the observed points P_j have positive components. Also, if m is the number of such components, then we shall also assume that there is at least one basis for m -dimensional space among the vectors P_j . We can then start by choosing P_0 as some point which is properly interior to the cone with apex at the origin and spanned by $P_1, \dots, P_n$. This can always be arranged since, for example,

$$P_0 = \sum_{j=1}^n P_j \text{ will have this property.}$$

An illustrative example may help to fix ideas and provide concrete insight for our procedures and interpretations. Thus, going back to the data of Figure 1 (without P_6) and choosing $P_0 = P_1$ initially, as before, we set up the simplex tableau for (14.2) and then proceed as in Table 1.

Notice that for (14.2) we always have a basis of natural slack unit vectors as in Stage 0. Notice further that the tableau entries under the structural vectors $P_1, \dots, P_5$ and the slack vectors e_1, e_2 are completely independent of the entries under the stipulations vector P_0 - i.e., the entries under the structural and slack vectors in each tableau are the same no matter what P_0 is selected.

In particular, consider Stage 2 in Table 1. As is evident from Figure 1, both P_3 and P_5 are efficient extreme points. Nevertheless, the facet spanned by them is not efficient. Correspondingly, the Stage 2 tableau is not an optimal tableau since $z_4 - c_4 = -\frac{1}{5} < 0$. But Stage 3* is an optimal tableau even though $z_0^* = 7/6 > 1$ tells us that P_1 is not efficient. Now replacing P_1 by $P_0 = P_4 v_4 + P_5 v_5$ for any $v_4, v_5 \geq 0$ and $v_4 + v_5 = 1$ will not change the $z_j - c_j$ but will give us a P_0 column of v_4, v_5 with $z_0^* = 1$ so the whole facet spanned by P_4, P_5 is efficient.

The lesson here is that we can start with P_0 an interior point and proceed through simplex tableaus to an optimal tableau. Then replacing P_0 by any convex combination of the optimal basis vectors we will achieve a $z_0^* = 1$. Hence we will have exhibited an efficient facet and, as we shall show, the equation of the hyperplane containing this facet will also then be available. For, as is well known,^{1/} the $z_j - c_j$ under the slack vectors in an optimal tableau are an optimal solution to the dual problem. I.e., these optimal $z_j - c_j$ values under the slack columns in the primal tableau are the optimal values $\omega = \omega^*$ for the dual problem,

$$\begin{aligned} \text{Min } g_0 &= \omega^T P_0 \\ (15) \quad &\text{with } \omega^T P_j \geq 1, j = 1, \dots, n \\ &\omega^T \geq 0 \end{aligned}$$

^{1/} See, e.g., [9] or [27].

and $\omega^{*T} P_i = 1$ for each P_i in an optimal basis. The superscript T represents transposition, as usual, so that, e.g., ω^T represents the transpose of the column vector ω with components $\omega_1, \dots, \omega_m$ and ω^{*T} denotes an optimum value for these variables in the above problem.

To show that ω^{*T} is orthogonal to the efficient facet spanned by these P_i , we need only show ω^* is orthogonal to any direction lying in the facet, e.g., to any vector which is the difference, $\bar{P} - \bar{\bar{P}}$, of two vectors in the facet. Since $\bar{P}$ and $\bar{\bar{P}}$ are in the facet,

$$\bar{P} = \sum_i P_i \bar{v}_i, \quad \bar{\bar{P}} = \sum_i P_i \bar{\bar{v}}_i$$

$$0 \leq \bar{v}_i, \bar{\bar{v}}_i \leq 1$$

(16.1) and

$$\sum_i \bar{v}_i = \sum_i \bar{\bar{v}}_i = 1$$

where summation is over the indexes indicated by these P_i . But then

$$\begin{aligned} \omega^{*T} (\bar{P} - \bar{\bar{P}}) &= \sum_i (\omega^{*T} P_i \bar{v}_i - \omega^{*T} P_i \bar{\bar{v}}_i) \\ (16.2) \quad &= \sum_i (1 \bar{v}_i - 1 \bar{\bar{v}}_i) = \sum_i \bar{v}_i - \sum_i \bar{\bar{v}}_i = 0, \end{aligned}$$

since $\sum_i \bar{v}_i = \sum_i \bar{\bar{v}}_i = 1$. Q.E.D. Hence, the ω^* corresponding to this efficient facet (and simplex) determined by this optimal basis is orthogonal (or normal) to it. Thus, ω^* is normal to the hyperplane containing this facet. The equation of this hyperplane is

$$(17) \quad \omega^{*T} x = 1$$

where x is any point in the linear space spanned by the totality of the P_j 's and e_i 's.

We now return to the two-dimensional example of Figure 1 in order to exemplify these developments. Taking Stage 3* in Table 1 for illustration, we have $\omega^*{}^T = (\frac{1}{3}, \frac{1}{6})$ as is apparent from the $z_j - c_j$ values listed under

the slack vectors at this stage. The points P_0 of the efficient facet are $P_0 = P_4 v_4 + P_5 v_5$ for all $v_4, v_5 \geq 0$ with $v_4 + v_5 = 1$. Thereby we have $\omega^*{}^T P_0 = (\omega^*{}^T P_4) v_4 + (\omega^*{}^T P_5) v_5$.

Moving to (17) we therefore have, for this two-dimensional case,

$$\omega^*{}^T x = \omega_1^* x_1 + \omega_2^* x_2 = 1$$

which, via the results available at Stage 3* in Table 1, become

$$\frac{1}{3} x_1 + \frac{1}{6} x_2 = 1.$$

In Figure 1, this facet is on a line with the segment connecting P_4 and P_5 constituting part of the efficient isoquant and, in fact, this segment may now be rendered explicitly as the set of points represented in

$$S [4, 5] \equiv \left\{ (x_1, x_2) : \frac{1}{3} x_1 + \frac{1}{6} x_2 = 1; 1 \leq x_1 \leq 2, 2 \leq x_2 \leq 4 \right\}.$$

Continuing to Stage 5* and observing the $z_j - c_j$ values under the slack at this stage we then obtain the expression for the adjacent isoquant

segment as

$$S [3, 4] \equiv \left\{ (x_1, x_2) : \frac{1}{6} x_1 + \frac{1}{3} x_2 = 1; 2 \leq x_1 \leq 4, 1 \leq x_2 \leq 2 \right\}.$$

Of course, the property to be emphasized is that the numerical values of the coefficients in these expressions are available without any extra effort.^{1/} Hence, also without extra effort, we have a new method of effecting numerical estimates of economic relations from observational data and, as remarked in our introduction, this should be of interest for any discipline (or use) where explicit numerical estimates of frontier relations from observational data are wanted.

Dual and Complementary Problems

For the single output case, we have now shown that the formulations (14) can be used to obtain the wanted measures of efficiency. Their duals provide the coefficients of the corresponding isoquant surface, facet by facet. Furthermore, by a suitable choice of solution method the wanted efficiency measures and coefficient estimates are simultaneously obtainable from either the primal or dual problems, as convenient, by solving only one of them. Finally, and fortunately, the readily available computer codes, which are generally built around the simplex and dual methods, have the desired properties.

Other topics such as the invariance of these efficiency measures are relegated to appendix for at least brief examination.

In this way we can continue toward extensions and generalizations

1/ The isoquant segment extending from P_3 to Q_1 may be represented by

$$s [3, 1') = \{ (x_1, x_2) : 1 = 0 x_1 + x_2 ; 4 \leq x_1, 1 \leq x_2 \}$$

or by

$$s [3, 1') = \{ (x_1, x_2) : 1 = \frac{1}{M} x_1 + x_2 ; 4 \leq x_1, 1 \leq x_2 \} ,$$

where the prime refers to the subscript on Q and the square bracket denotes inclusion of the point whose subscript is next to the bracket, while the round bracket denotes exclusion. The expressions refer to the segments from P_3 to Q_1 in Figure 1, according to whether (11.2) or (14.2) is employed - and a similar development may obviously be undertaken for the segment extending from P_5 to Q_1 .

in the next section. First, however, we try to relate what we have already done to other streams of research in the pure theory of production. In particular we shall try to relate it to other duality approaches to which have followed on the original contributions of Samuelson [30] and Shepherd [34] and [35], and which have recently been given a detailed review in Diewert [22.2] and the commentaries accompanying it.

As Lau notes on p. 178 of one of these commentaries [22.1], a variety of approaches toward duality in the theory of production have been essayed which, in turn, can also be related to various parts of mathematics that range from the classical calculus of variations (e.g., duality such as in the classical isoperimetric problem) to modern mathematical programming and game theory (e.g., the Kuhn-Tucker theorem and related topics).

Our approach has evidently used the relations of equality and inequality from linear programming and hence is in the latter tradition. Other approaches involve functional characterizations via pairs of problems in which optimizations applied to a given functional form in one problem are used to obtain the form of the other function. For instance, cf., e.g., Diewert [22.2], pp. 108-9, a given production function is used to obtain a corresponding cost function in explicit form from a constrained minimization problem. Alternatively, availability of the cost function makes it possible to obtain the production function via a related problem of constrained maximization.

In a usage which evidently differs from the one we have been employing^{1/}, this is said to involve a relation of duality between the indicated cost and production function problems. See, e.g., Diewert [22.2], p.108. For our own immediate purposes, however, we shall refer to these as complementary problems in order to avoid possible confusion in terminology.

Thus, consider the problem

$$\begin{aligned}
 & \min. \quad c^T x \\
 & \text{with} \quad a_s^T x \geq y, \quad s=1, \dots, k \\
 (18.1) \quad & -P\lambda + I x = 0 \\
 & \lambda \geq 0,
 \end{aligned}$$

wherein P represents a matrix of m -component vectors P_j , $j=1, \dots, n$, of the observational variety described in the preceding section, say, and I is the corresponding identity matrix. Here we are stating a decision problem involving a choice of input vectors x which lie in the cone that can be formed from these P_j via the vector of non-negative variables λ . We are also using $a_s \equiv \omega_s^*$ for the s^{th} vector of coefficients for the corresponding $s=1, \dots, k$ efficient facets determined from observational data in the manner indicated in the preceding section. Hence the wanted production function is at hand for obtaining any prescribed output level y under the minimization associated with any vector of positive constants c^T .

^{1/} For instance, the resulting cost and production functions are in different dimensions and hence cannot have relations of equality and inequality between them without further transformations and development. See, e.g. [22.3].

I.e., c^T is a vector with constants $c_i > 0$, $i=1, \dots, m$, for its components.

As already indicated, we refer to (18.1) as a problem complementary to the one by which we determine these a_s values. Then we refer to the following as its dual,

$$\begin{aligned}
 (18.2) \quad & \max \quad y \sum_{s=1}^k \eta_s \\
 & \text{with} \quad \sum_{s=1}^k \eta_s a_s^T + u^T I = c^T \\
 & \quad \quad \quad - u^T P = 0 \\
 & \quad \quad \quad \eta_s \geq 0, \quad s=1, \dots, k.
 \end{aligned}$$

Via the duality theory of mathematical programming we have

$$(19.1) \quad c^T x \leq y \sum_{s=1}^k \eta_s$$

for all x and η_s which satisfy the constraints and

$$(19.2) \quad c^T x^* = y \sum_{s=1}^k \eta_s^*$$

at an optimum.

If we now interpret each c_i in c^T as the price (per unit) of a particular input then parametric variation of these components may be employed to develop the corresponding factor demands. The expression on the right is then the resulting (dual) cost function. I.e., the latter is a true economic (opportunity) cost function in the sense that the amount of output y will be produced at a loss unless an amount of at least $\sum \eta_s^*$ per unit can be obtained from its sale.

In fact, if wanted, the corresponding profit function

$$\pi(y) = R(y) - C(y)$$

can also be developed wherein the revenue function $R(y) = p y$ and the cost function $y \sum \eta_s^*$ determines whether $\pi(y)$ is positive or negative for each value of y . Of course, if these functions are both linear and homogeneous then under the usual assumptions of profitmaximizing behaviour, the output choices will be zero or infinite. On the other hand factor supply functions and product demand functions may be introduced which are piecewise linear, say, and still other devices may be employed to circumvent these difficulties.

It is not our intention to undertake developments of the latter variety here. Rather our purpose is only to note the possibility of joining one stream of production function research, motivated largely by considerations of empirical (e.g., frontier)

estimation^{1/} and policy analysis^{2/} to another stream motivated largely by considerations of theoretical (and pedagogical) clarification. Note, for instance, that we might also have proceeded via cost (or factor price) data as a basis for our empirical estimates. Then, as the preceding development suggests, we could have moved into a set of complementary dual problems to obtain the corresponding production function.

^{1/} See Aigner, Amemiya and Poirier [2] for an up-to-date discussion.

^{2/} See, e.g., Carlsson [6] for a use of Farrell efficiency measures to assess, by reference to Leibenstein's concept of X-efficiency -- see [25] and [26] -- whether Swedish manufacturing firms were operating at or near the frontier of their possible (relative) efficiency.

In this paper, however, we shall continue to employ an approach which does not make direct use of price or cost data.^{1/} Partly this is because we want to be able to apply our results to public institutions or services where price and cost data are difficult to come by. Partly also we share the concern expressed by Lau [22.1] p.196, about the relative ranges of error and variation in price (and cost) as compared to quantity data.

^{1/}Note should probably be made of Farrell's own use of cost data to deal with increasing returns to scale problems in applying his efficiency measure to industrial data. See e.g., Farrell and Fieldhouse [19]. See also [33].

Extension to Multiple Outputs

Following the just indicated lines, we now turn to the case of multiple outputs and write

$$\begin{aligned}
 & \max z_o \\
 & \text{subject to} \\
 (20) \quad & - \sum_{j=1}^n y_{rj} \lambda_j + y_{ro} z_o \leq 0, \quad r = 1, \dots, s \\
 & \sum_{j=1}^n x_{ij} \lambda_j \leq x_{io}, \quad i = 1, \dots, m \\
 & \lambda_j \geq 0 \quad j = 1, \dots, n
 \end{aligned}$$

and utilize the optimal $z_o = z_o^*$ for the wanted scalar measure of efficiency. That is, z_o^* is to play the same role here as in (13), above.

In fact, these same conditions for efficiency will continue to apply except that now the points P_j consist of output vectors Y_j with components y_{rj} as well as input vectors X_j with components x_{ij} . For example, the previous normed observations are replaced by new

$$(21) \quad P_j = \begin{pmatrix} Y_j \\ X_j \end{pmatrix}, \quad j=1, \dots, n,$$

in which the outputs y_{rj} and inputs x_{ij} are separately displayed (in the same position) for each "firm".

The previous interpretations also extend in a very straightforward manner to the present (multiple output) case. Thus, e.g., the appearance

of positive slack in an optimal tableau for some $P_k = P_o$ now can mean that some output can be increased without requiring more of any input, and, of course, it can also mean (as before) that some inputs can be decreased without reducing any outputs.

Similarly $z_o^* \geq 1$ also continues to apply wherein $z_o^* > 1$ now means that it is possible to expand all of the $r = 1, \dots, s$ outputs by moving along the indicated ray and to do so, moreover, without requiring further inputs in any of these $i = 1, \dots, m$ categories.

The dual to (20) is,

$$\begin{aligned} \text{Min } g_o &= \sum_{i=1}^m \omega_i x_{io} \\ \text{subject to} \\ (22) \quad & - \sum_{r=1}^s \mu_r y_{rj} + \sum_{i=1}^m \omega_i x_{ij} \geq 0 \\ & \sum_{r=1}^s \mu_r y_{ro} = 1 \\ & \mu_r, \omega_i \geq 0 \end{aligned}$$

This problem is to be formed in principle by substituting x_{ik}, y_{rk} for the x_{io}, y_{ro} above. In practice, far fewer substitutions will be necessary since many firms will be recognized as efficient (and inefficient) from already achieved optimal tableaux. See Table 1, ff. ^{1/}

^{1/} The designation of efficient bases continues as before but continuation by the dual method is more complex and may require recourse to other routines. See, e.g., [11].

Of course, the concept of an isoquant becomes ambiguous when more than a single output is being considered and so we replace it by other economic concepts such as "production possibility sets" and the "activities" to which they are related.^{1/} The duality relations of linear programming continue to apply to the above formulations, however, and so with a suitable choice of methods (any adjacent extreme point method will do) we can again obtain the values of the input and output coefficients of the corresponding activity vectors. From these expressions the corresponding efficiency (frontier) surfaces along with all of the other production possibilities (including those below the frontiers) can then also be obtained, and so on.

We can now also secure further clarification for all of the preceding developments by introducing

$$(23) \quad w^* = (\mu^*, w^*)$$

wherein μ^* represents the vector with components μ_p^* for the output coefficients and w^* has components w_i^* for the input coefficients. In general, some of these coefficients will be zero and, of course, we must also expect the possibility of alternate optima whereupon the efficient surface degenerates to a lower dimension. Special interest attaches to the case in which none of the components of w^* are zero, if only because this has been customarily assumed for the study of substitution as well as expansion and contraction possibilities in economics. Hence, we may observe that a

^{1/} Cf., e.g., K. J. Arrow and F. H. Hahn [4] pp. 52 ff.

straightforward consideration of the theory underlying the simplex method^{1/} shows that this will occur -- i.e., all components of w^* in every optimum will be positive -- if and only if there are no slack vectors in any optimal basis. This is a natural requirement since, from an economics standpoint, these vectors represent a computing artifact and are not a part of the "economically admissible" observations.

We now proceed to try for still further clarification from (23) by drawing on the theory of "fractional programming"^{2/} from which we introduce the transformations

$$(24) \quad \begin{aligned} w_i &= t v_i \\ u_r &= t u_r \end{aligned}$$

where, as before, $i = 1, \dots, m$ and $r = 1, \dots, s$, which, with $t > 0$, implies $v_i, u_r \geq 0$ whenever $w_i, u_r \geq 0$. Because of the condition $\sum_{r=1}^s u_r y_{ro} = 1$ in (22), one can recognize -- cf [10] -- that (22) is equivalent to the following linear fractional programming problem

$$(25) \quad \begin{aligned} \min f_o &= \frac{\sum_{i=1}^m v_i x_{io}}{\sum_{r=1}^s u_r y_{ro}} \\ \text{subject to} \\ \sum_{i=1}^m v_i x_{ij} - \sum_{r=1}^s u_r y_{rj} &\geq 0, \quad j = 1, \dots, n \\ v_i, u_r &\geq 0. \end{aligned}$$

^{1/} See, e.g., [9] or [17].

^{2/} See [10] and [7].

Further this linear fractional problem is evidently equivalent to the following nonlinear ratio formulation

$$\begin{aligned}
 (26) \quad \min f_o &= \frac{\sum_{i=1}^m v_i x_{io}}{\sum_{r=1}^s u_r y_{ro}} \\
 &\text{subject to} \\
 &\frac{\sum_{i=1}^m v_i x_{ij}}{\sum_{r=1}^s u_r y_{rj}} \geq 1, \quad j=1, \dots, n \\
 &v_i, u_r \geq 0.
 \end{aligned}$$

Equivalently, we can also replace this with

$$\begin{aligned}
 (27) \quad \max h_o &= \frac{\sum_{r=1}^s u_r y_{ro}}{\sum_{i=1}^m v_i x_{io}} \\
 &\text{subject to} \\
 &\frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1, \quad j=1, \dots, n \\
 &u_r, v_i \geq 0.
 \end{aligned}$$

Thus, we can interpret our efficiency characterization in (26) as finding a minimum ratio of weighted inputs to outputs for weights such that the resulting ratio for every observed firm is at least unity. This means that no firm can be rated as more than 100% efficient in such a system. See (27).

We can obtain further clarification via the simpler single output case. The latter can also be rendered in the form (14) wherein we can also eliminate the subscript r , which is no longer necessary, and write

$$\begin{aligned}
 & \max y_0 z_0 & \min g_0 = \sum_{i=1}^m \omega_i x_{i0} \\
 & \text{subject to} & \text{subject to} \\
 & - \sum_{j=1}^n y_j \lambda_j + y_0 z_0 \leq 0 & \sum_{i=1}^m \omega_i x_{ij} - \mu y_j \geq 0 \\
 (28) & & \mu = 1 \\
 & \sum_{j=1}^n x_{ij} \lambda_j \leq 0 & \\
 & \lambda_j \geq 0 & \omega_i \geq 0 \\
 & & i=1, \dots, m \\
 & & j=1, \dots, n
 \end{aligned}$$

in place of (20) and (22).

We have altered the functional (but in an inessential way) in moving from (20) and (22) by multiplying z_0 by the positive constant y_0 , the single output quantity, for each $P_k = P_0$. In this way we replace the isoquant with the corresponding production (function) surface by means of the relation

$$(29) \quad y_0 z_0^* = \sum_{i=1}^m \omega_i^* x_{i0}$$

which holds at an optimum. Because z_0 is in abstract measure (i.e., it

is dimensionless), it follows that w_i^* is in units of output for the i^{th} input, as required for a production function coefficient. Moreover, $z_0^* \geq 1$ with $z_0^* > 1$ mapping the observed output up onto the production surface and hence

$$(30) \quad y = \sum_{i=1}^m w_i^* x_i$$

is the explicit expression for this portion of the production function surface. The development is otherwise the same as was given earlier for the isoquant expressions.

We need to underscore, perhaps, that this production function, and hence the preceding isoquant expressions, differs from the ones usually used in two ways. First it is oriented toward individual decision making units (e.g., a firm) rather than industrywide (or even broader) aggregates. Second, it adjusts the empirical observations up to the efficiency surface rather than simply fitting an average relation by ordinary (e.g., least squares) statistical methods.^{1/} That is, it adjusts each input for any slack that may be present and then multiplies the observed y_0 by $z_0^* \geq 1$ to reach the efficient production function surface with which it is associated in (30).

We might argue that this approach is more firmly fixed in the economic theory of production, as usually developed in terms of an assumed optimizing behavior of individual decision making units. Hence, it can shed light on a variety of theoretical issues (as well as issues of policy) such as, e.g., "The Xistence of X-Efficiency", which has come under critical challenge in G. J. Stigler's discussion^{2/} of a well known article by H. Leibenstein.^{3/}

^{1/} See, however, the attempt by Afriat [1] to justify a use of least squares for production function estimation.

^{2/} See [37]. See also [6] and [21].

^{3/} See [25] and [26].

Moreover, the approach we are suggesting can be used to derive wanted aggregates of inputs and outputs -- i.e., an "aggregate production function" -- and furthermore, it permits an additional choice between the y_0 values only (i.e., using the original output observations) or a $y_0 z_0^*$ (i.e., using efficiency adjusted output observations) according to the purpose being served. In choosing among the many possibilities for defining such aggregate functions, we might provide some guidance as follows. The "representative" efficient surface (e.g., the unit isoquant in the single output case) consists of facets, each of which is a simplex. Each such simplex spans a convex cone. Thus when considering the aggregation, i.e., the summing, of two or more inputs, the related efficient aggregate output will be the sum (or aggregate) of the efficient outputs of the individual firms as long as they are in the same cone. If the individual inputs are from different cones, the aggregate input may lie in any one of the individual cones and it can lie in other cones as well. In this case (i.e., at least two inputs from different cones), the efficient aggregate output will exceed the sum of the efficient individual outputs by virtue of the strict sub-additivity ^{1/} of our efficient production function for input vectors from different cones.

Other relations of disaggregation and aggregation may also be employed, of course, according to purposes that might be served thereby. Inter alia, none of the above is intended to imply that customary approaches

^{1/} See [9] pp. 284 ff.

to the study of empirical production relations must necessarily be abandoned.^{1/}
It means rather that we now have numerous other alternatives that can be shaped and applied for different purposes as the contexts may suggest.

Data at "individual firm" levels may not be available. Indeed this is characteristic of many statistical collections (e.g., the U. S. Census of Manufactures)^{2/} of private enterprise data. Such data are more likely to be available for many public service organizations and organizations in regulated industries. Thus, as already noted, we have oriented the above development in ways which are intended for use in the latter cases -- rather than the former, where price/profit data are likely to provide efficiency measures which evaluate "direction" as well as "magnitude" for optimizing choices of outputs and inputs, and where a variety of measures (e.g., enforcement of market competition) can be employed to ensure that the indicated optimizations are actually undertaken. See, e.g. [36].

1/ Recent years have seen a considerable amount of attention focussed on methods for estimating efficiency frontiers and a review of some of that literature may be found in [2]. The conditions for aggregation from individual to industry (or economy-wide) total relations may also be fulfilled in a variety of ways. For instance it is shown in [14] that the aggregates should take the form of "extended Cobb-Douglas functions", with output elasticities summing to unity, whenever all firms (a) have "smooth" homogeneous production functions, of possibly varying degree and (b) engage in cost minimizing behavior with common (fixed) input prices. See also [13].

2/ Such data might also become regularly available on an individual enterprise basis if the SEC pushes further in its current emphasis on "line of business" reporting, etc.

Summary and Conclusion

Still further extensions might be essayed. One such possibility would include various "weighted" measures of efficiency, etc., to comprehend situations such as those in which a single decision maker is to be accommodated. Coupled with suitable computation routines, this could be used to illuminate alternative choices (and their associated weights) for multiple output situations, say, in which some outputs might be decreased in order to secure increases in other outputs.^{1/}

Another possibility might cover extensions for increasing returns to scale and a variety of nonlinear interactions as well.^{2/} Alternatively a variety of constrained optimization models might also be erected not only for statistical estimation but also for studying allocation and reallocation possibilities under some of the different kinds of production functions discussed in the preceding section.

Finally, we come to other questions such as computation and statistical considerations. First we turn to computation. In most cases there will be relatively few constraints and this should cause no great trouble since available linear programming codes can readily deal with problems involving virtually unlimited numbers of variables, provided the constraints are few in number. Of course when one wants to expand

^{1/} See, e.g., the discussion of the spiral method and its illustrative application to such a weighted efficiency model in [9].

^{2/} See, e.g., the discussions in [19] and [20].

the inputs to accomodate different types of labor, say, or different vintages of capital then the constraints will increase in number. The same is true for other extensions of production function concepts, e.g., to deal with quality-of-life estimates in the form of multiple outputs as N. Terleckyj has done in [38].

Research pointed toward new and improved algorithms may well become very worthwhile as such extensions and their related applications proceed. The same is true for statistical characterizations that might be employed to augment the approaches used in this paper. Here we might note, for example, that Aigner and Chu in [3] have already suggested a use of chance constrained programming and that Timmer [39] has experimented with such approaches, albeit in only an embryonic way. For our own part, we prefer to leave such developments until a variety of applications makes it apparent how the appropriate models (and methods) might best be undertaken. That, for example, is the way in which chance-constrained programming was first invented.^{1/} At this point we might then close by noting that we ought to give primacy to the concepts of economics (as Farrell has done) at least until we can see that these are clearly overpowered by statistics -- e.g., by virtue of data or other considerations -- in which case we, perhaps, ought to consider revising our underlying economic concepts and definitions along lines which differ from the versions which are embodied in the present paper.

^{1/}See [15] and [8].

Appendix

In all of Farrell's treatment an important consideration is the convexity of the set of production possibilities. There may, of course, be many different measures for the amount of inputs or outputs and so we need to consider how the resulting measure of efficiency may be altered as these input and output measures are altered.

For clarity we consider the single output case. There may even be transformations of the $n+1$ coordinates of inputs plus outputs into other coordinate systems (with $n+1$ coordinates) in which, say, a nonconvex set of production possibilities is transformed into a convex set -- as required for Farrell's analysis. Provided that such a transformation is "bicontinuous",^{1/} however, boundary points remain boundary points and interior points remain interior points. Thus when there are two different coordinate representations both of which have convex sets of production possibilities, efficient points remain as such in both systems and the same is true for nonefficient points. Hence the characterization of any point as efficient or non-efficient will remain invariant under any such transformation. The numerical magnitude may, of course, change. Specific examples can be given, but it is important to note that even this numerical value remains invariant under any "homothetic" transformation. The latter, which refers to constant changes of scale, is the one that will usually be of interest and hence, in this respect, too, Farrell has provided an important opening for empirical work and related policy applications which our models also preserve.

^{1/} See, e.g., J. L. Kelley [24].

Bibliography

- [1] Afriat, S.N., "Efficiency Estimation of Production Functions," International Economic Review, 13, No. 3, Oct., 1972, pp. 568-583.
- [2] Aigner, D.J., T. Amemiya, P. J. Poirier, "On the Estimation of Production Frontiers: Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function," International Economic Review, XVII, No. 2, June 1976, pp. 377-396.
- [3] Aigner, D.J. and S.F. Chu, "On Estimating the Industry Production Function," American Economic Review, LVIII, Sept. 1968, pp. 826-839.
- [4] Arrow, K.J. and F.H. Hahn, General Competitive Analysis, (San Francisco: Holden-Day, Inc. 1971)
- [5] Bector, C.R., "Duality in Linear Fractional Programming," Utilitas Mathematica, 4, 1973, pp. 155-168.
- [6] Carlsson, B, "The Measurement of Efficiency in Production: An Application to Swedish Manufacturing Industries, 1968" Swedish Journal of Economics, 1972.
- [7] Charnes, A. and W.W. Cooper, "An Explicit General Solution in Linear Fractional Programming," Naval Research Logistics Quarterly, 20, No. 3, Sept. 1973.
- [8] Charnes, A. and W.W. Cooper, "Chance Constrained Programming," Management Science, Oct. 1959.
- [9] Charnes, A. and W.W. Cooper, Management Models and Industrial Applications of Linear Programming, Vol. I, (New York: John Wiley and Sons, Inc. 1961).
- [10] Charnes, A. and W.W. Cooper, "Programming With Linear Fractional Functionals," Naval Research Logistics Quarterly, 9, Nos. 3 and 4, Sept.-Dec., 1962, pp. 181-185.
- [11] Charnes, A. and W.W. Cooper, "Structural Sensitivity Analysis and an Exact Product Form Left Inverse," Naval Research Logistics Quarterly, Vol. 15, Dec., 1968, pp. 517-521.
- [12] Charnes, A., W.W. Cooper and E.L. Rhodes, "Expositions, Interpretations and Extensions of Farrell Efficiency Measures," Research Report (Pittsburgh: Carnegie-Mellon University School of Urban and Public Affairs, April 30, 1975.)
- [13] Charnes, A., W.W. Cooper and A. Schinnar, "A Theorem on Homogeneous Functions and Extended Cobb-Douglas Forms," Proceedings of the National Academy of Sciences, Oct., 1976, Vol. 73, No. 10.
- [14] Charnes, A., W.W. Cooper, and A.P. Schinnar, "An Extended Cobb-Douglas Form for Use in Production Economics." Pittsburgh, Pa. Carnegie-Mellon University, School of Urban and Public Affairs, Institute of Physical Planning, Research Report (mimeo). Dec. 1975.

- [15] Charnes, A., W.W. Cooper and G.H. Symonds, "Cost Horizons and Certainty Equivalents: An Approach to Stochastic Programming of Heating Oil by the Horizon Method," Management Science, April 1958.
- [16] Cooper, W.W., "A Proposal for Extending the Theory of the Firm," Quarterly Journal of Economics LXV, NO.1, Feb. 1951.
- [17] Dantzig, G.B., Linear Inequalities and Related Systems (Princeton; Princeton University Press, 1963).
- [18] Farrell, M.J., "The Measurement of Productive Efficiency," Journal of the Royal Statistical Society, Series A, Part III, 1957, pp. 253-290.
- [19] Farrell, M.J. and M. Fieldhouse, "Estimating Efficient Production Functions under Increasing Returns to Scale," Journal of the Royal Statistical Society, Series A, Part II, 1962, pp. 252-267.
- [20] Førsund, F.R., Studies in the Neo-Classical Theory of Production (Oslo: University of Oslo Institute of Economics Memorandum, 1974. See especially Ch. II: Measures of Efficiency of Production Units" and Ch. IV "On the Measurement of Productive Efficiency").
- [21] Førsund, F.R. and L. Hjalmarsson, "Comment on Bo Carlsson's 'The Measurement of Efficiency in Production: An Application to Swedish Manufacturing Industries, 1968,'" Swedish Journal of Economics, 1974.
- [22] Intrilligator, M.D. and D.A. Kendrick, eds., Frontiers of Quantitative Economics, Vol. II. (Amsterdam: North Holland Publishing Co., 1974).
 - [22.1] Lau, L.J., "Comments" pp. 176-199.
 - [22.2] Diewert, W.E., Chap. 3, "Applications of Duality Theory," pp. 106-171.
 - [22.3] Shephard, R.E., "Comments" pp. 200-206.
- [23] Jaganathan, R., "Duality for Nonlinear Fractional Programs," Zeitschrift für Operations Research, Band 17, Seite 1-3, Würzburg: Physica-Verlag, 1973.
- [24] Kelley, J.L., General Topology (New York: Van Nostrand, 1955).
- [25] Leibenstein, H., "Allocative Efficiency vs. X-Efficiency," The American Economic Review, June 1966, pp. 392-415.
- [26] Leibenstein, H., Beyond Economic Man (Cambridge: Harvard University Press, 1976).
- [27] Lemke, C.E., "The Dual Method of Solving the Linear Programming Problem," Naval Research Logistics Quarterly, Vol. I, 1954, pp. 36-47.

- [28] Motzkin, T.S., H. Raiffa, G.L. Thompson and R.M. Thrall, "The Double Description Method," in Contributions to the Theory of Games, Vol. II, Annals of Mathematics Studies, 24, Princeton University Press, 1953.
- [29] Robinson, A., Non-Standard Analysis, Studies in Logic and the Foundations of Mathematics (Amsterdam: North Holland Press, 1966).
- [30] Samuelson, P.A., "Prices of Factors and Goods in General Equilibrium," The Review of Economic Studies, 21, 1953-1954, pp. 1-20.
- [31] Sato, K., Production Functions and Aggregation (Amsterdam: North Holland Publishing Co., 1975).
- [32] Schaible, S., "Parameter-free Convex Equivalent and Dual Programs of Fractional Programming Problems," Zeitschrift für Operations Research, Band 18, Seite 187-196, Würzburg: Physica-Verlag, 1974.
- [33] Seitz, W.D., "Productive Efficiency in the Steam-Electric Generating Industry", Journal of Political Economy, Vol. 79, No. 4, 1971, pp. 878-886.
- [34] Shephard, R.W., Cost and Production Functions (Princeton: Princeton University Press, 1953).
- [35] Shephard, R.W., The Theory of Cost and Production Functions (Princeton: Princeton University Press, 1970).
- [36] Stigler, G.J., The Organization of Industry (Homewood, Illinois: Richard D. Irwin, Inc., 1968).
- [37] Stigler, G.J., "The Existence of X-Efficiency," The American Economic Review, March 1976, pp. 213-216.
- [38] Terleckyj, N.E., Improvements in the Quality of Life (Washington, D.C.: The National Planning Association, 1975).
- [39] Timmer, C.P., "Using a Probabilistic Frontier Production Function to Measure Technical Efficiency," Journal of Political Economy 79 (4), 1971, pp. 776-794.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Center for Cybernetic Studies ✓ The University of Texas		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP	
3. REPORT TITLE ⑥ Measuring the Efficiency of Decision Making Units with Some New Production Functions and Estimation Methods.			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates)			
5. AUTHOR(S) (First name, middle initial, last name) ⑩ A./Charnes W. W./Cooper E./Rhodes		⑨ Research rept.,	
6. REPORT DATE October 8, 1976		7a. TOTAL NO. OF PAGES 50	7b. NO. OF REFS 39
8a. CONTRACT OR GRANT NO. ⑮ N00014-75-C-0616 and N00014-76-C-0932		9a. ORIGINATOR'S REPORT NUMBER(S) Center for Cybernetic Studies Research Report CCS-276	
8b. PROJECT NO. NR047-021		⑭	
c. ⑪ 8 Oct 76		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) ⑫ 54p.	
d.			
10. DISTRIBUTION STATEMENT This document has been approved for public release and sale; its distribution is unlimited.			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY Office of Naval Research (Code 434) Washington, D.C.	
13. ABSTRACT A new series of linear programming models is used to clarify and extend a measure of efficiency introduced by M. J. Farrell. Multiple output as well as single output versions are provided which relate this efficiency to ordinary Pareto-Koopmans optimality conditions. Their duals supply numerical estimates of production coefficients. This duality, from mathematical programming, is distinguished from another in the theory of production, (as found in the works of Samuelson, Shephard, et al) and then used to provide new nonlinear duals which open contacts with other mathematical programming developments in areas like fractional programming, etc.			

Unclassified

Security Classification

A-31408

1B

