Measuring the efficiency of decision making units

A. CHARNES
The University of Texas, Austin TX, U.S.A.

W.W. COOPER
Graduate School of Business Administration, Harvard University, Cambridge MA, U.S.A.

E. RHODES
State University of New York at Buffalo, Albany NY, U.S.A.

A nonlinear (nonconvex) programming model provides a new definition of efficiency for use in evaluating activities of not–for–profit entities participating in public programs. A scalar measure of the efficiency of each participating unit is thereby provided, along with methods for objectively determining weights by reference to the observational data for the multiple outputs and multiple inputs that characterize such programs. Equivalences are established to *ordinary* linear programming models for effecting computations. The duals to these linear programming models provide a new way for estimating extremal relations from observational data. Connections between engineering and economic approaches to efficiency are delineated along with new interpretations and ways of using them in evaluating and controlling managerial behavior in public programs.

This paper represents a revision of earlier versions entitled, respectively, 'Expositions, Interpretations, and Extensions of the Farrell Efficiency Measure', and 'Measuring the Efficiency of Decision Making Units With Some New Production Functions and Estimation Methods'. See [11], [12] and also [9]. This research was partly supported by NSF Grant No. SOC76-15876 'Collaborative Research on the Analytical Capabilities of a Goals Accounting System'. It is also part of the program of the Navy Personnel Research and Development Center and is partly supported by Project NR 047-021, 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 School of Urban and Public Affairs.

European Journal of Operational Research 2 (1978) 429–444.

1. Introduction

This paper is concerned with developing measures of 'decision making efficiency' with special reference to possible use in evaluating public programs. As we shall use it, the term 'program' will refer to a collection of decision making units (DMU's) with common inputs and outputs. These outputs and inputs will usually be multiple in character and may also assume a variety of forms which admit of only ordinal measurements. For example, in an educational program like 'Follow Through' [1], the efficiency of various schools, viewed as DMU's in this program, may be measured by reference to outputs involving the standard education categories: viz., cognitive, affective and psycho-motor skills via, respectively, (1) arithmetic scores, (2) psychological tests of student attitudes, e.g., toward the community and (3) student ability to understand and control bodily motions, e.g., by observing their ability to tread water and turn from front to back (and vice versa) in a swimming pool [2]. These are all to be regarded as 'valued' outputs even when there is no apparent market for them or even when other possible sources for reasonably supportable systems of weights are not readily available. The inputs may similarly range from fairly easy to measure (and weight) quantities like 'number of teacher hours' and extend to more difficult ones line 'time spent in program activities by community leaders and/or parents'.

Our use of terms like 'DMU' (decision making unit) and 'programs' will help to emphasize that our interest is centered on decision making by not-for-profit entities rather than the more customary 'firms' and 'industries'. It will also help us to emphasize that our data (as in the above example) are not readily weighted by reference to market prices [3] and/or other economic *desiderata* – such as costs of producing income earning capacity in students, with related rates of discount – in accordance with the

[1] A discussion of this Federally sponsored program which includes a use of the efficiency measures we shall be discussing may be found in [21]. This includes a use of various statistical tests of significance (using the so-called Kullback-Leibler statistic), which will not be discussed in the present paper.

[2] See the discussion in [3] for a use of measures like these in program- planning- budgeting (PPBS) contexts.

[3] We are referring to *actual* market prices and costs. Later we shall show how to obtain estimates of (optimal) production coefficients and relate them to theoretical (opportunity) costs and prices.

ways in which some public sector activities are sometimes evaluated.

Naturally we shall want to relate our ideas to developments in economics. This will be done by reference to production functions and related concepts such as 'cost duality', etc. Although adaptations of these concepts will be needed we shall also try to indicate what is involved at suitable points in this paper. (See below, Section 6, for instance.)

We shall also want to relate our ideas to other disciplines, like engineering, which are also concerned with efficiency measurement. This will be done not only in the interests of greater unity but also in the interest of distinguishing between efficiencies associated with an underlying production 'technology' and those due to managerial decision making when the former can be identified and separated from the latter by, e.g., engineering characterizations.

Of course, when this cannot be done (the usual case in empirical economics) [4] we will need to rest content with the somewhat less satisfactory concept of 'relative efficiency'. The latter will be determined by reference to suitably arranged 'rankings' of the observed results of decision making by various DMU's in the same program (e.g., the different schools in program Follow Through) while allowing for the fact that different amounts of inputs (sometimes legally stipulated) may be involved so that, e.g., some DMU's are more like members of one subset and less like members of other subsets, etc., in the 'amounts' of particular inputs and outputs utilized.

The meaning and significance to be accorded these characterizations will be clarified in the sections that follow. First we shall introduce our proposed measures and models. Then we shall provide characterizations which are wholly computational. Relations to selected lines of ongoing research will be delineated, followed by methods of estimation and interpretation in terms of simple numerical illustrations and analytical characterizations. A concluding section will then summarize what has been done and point up relevant shortcomings along with possible further lines of development.

[4] Such separation is even more difficult in public sector programs such as education, public safety, etc., where the meaning of a 'technology' is likely to be more ambiguous than in the case of manufacturing in the private sector, and even many service operations.

2. Model and definition

Our proposed measure of the efficiency of any DMU is obtained as the maximum of a ratio of weighted outputs to weighted inputs subject to the condition that the similar ratios for every DMU be less than or equal to unity. In more precise form,

$$\max h_0 = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}} \tag{1}$$

subject to:

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leqslant 1 ; \qquad j = 1, ..., n,$$

$$u_r, v_i \geqslant 0 ; \quad r = 1, ..., s ; \quad i = 1, ..., m.$$

Here the y_{rj}, x_{ij} (all positive) are the known outputs and inputs of the jth DMU and the $u_r, v_i \geqslant 0$ are the variable weights to be determined by the solution of this problem – e.g., by the data on *all* of the DMU's which are being used as a reference set. The efficiency of one member of this reference set of $j = 1, ..., n$ DMU's is to be rated relative to the others. It is therefore represented in the functional, for optimization – as well as in the constraints – and further distinguished by assigning it the subscript '0' in the functional (but preserving its original subscript in the constraints). The indicated maximization then accords this DMU the most favorable weighting that the constraints allow.

For the DMU's which concern us, these x_{ij} and y_{rj} values, which are constants, will usually be observations from past decisions on inputs and the outputs that resulted therefrom. We can, however, replace some or all of these observations by theoretically determined values if we wish (and are able) to conduct our efficiency evaluations in that manner.

Consider, for instance, the following definition (quoted from [14]) from the field of combustion engineering – viz., 'efficiency is the ratio of the actual amount of heat liberated in a given device to the maximum amount which could be liberated by the fuel [being used]'. In symbols,

$E_r = y_r/y_R$

where

y_R = Maximum heat that can be obtained from a given input of fuel,

y_r = Heat obtained by the input being rated from the same fuel input.

Although the definition of efficiency varies from one engineering field to another, the one above captures the essentials – viz., the rating is relative to some maximum possibility so that, always, $0 \leqslant E_r \leqslant 1$.

We can also obtain the above defined E_r from (1) as follows. For any given input amount x substitution in (1) gives

$$\max h_0 = \frac{uy_0}{vx_0}$$

$$\text{s.t.} \quad \frac{uy_R}{vx_R} \leqslant 1 , \quad \frac{uy_r}{vx_r} \leqslant 1 , \quad u, v \geqslant 0,$$

where $r = 0$ in the functional designates that the latter is being rated.

Let u^*, v^* represent an optimal pair of values. Since $y_R \geqslant y_r$ and $x_R = x_r = x$ this implies $u^*y_R = v^*x_R$ and using $x_0 = x$ we then have the functional equal to y_r/y_R as required.

In common with most engineering definitions we have here confined our development to ratios of single ourputs and/or inputs, or weighted sums thereof. The latter may be determined, again by engineering considerations (e.g., efficient fuel combinations), which are ordinarily not available for the economic applications we are considering. Provided we have the indicated observations on inputs and outputs for individual DMU's, however, we can at least achieve 'relative efficiency' ratings along the lines that we have been suggesting. This is the way the rest of the paper will be developed although, as already indicated, we can also insert engineering or other data for such ratings, if we wish, in various combinations.

Note that our weightings, as above, are objectively determined to obtain a (dimensionless) scalar measure of efficiency in any case [5]. I.e., the choice of weights is determined directly from observational data subject only to the constraints set forth in (1). Under these observations and constraints no other set of common weights will give a more favorable rating relative to the reference set. Hence if a (relative) efficiency rating of 100% is not attained under this set of weights then it will also not be attained from any other set.

[5] Scaling and invariance properties which are dealt with in [12] – see also [21] – will not be discussed in this paper.

3. Reduction to linear programming forms

The above model is an extended nonlinear programming formulation of an ordinary fractional programming problem. We have elsewhere (in [10] and [7]) supplied a complete theory in terms of which fractional programming problems may be replaced with linear programming equivalents. We therefore propose to use that theory here to make the above formulation computationally tractable for the large numbers $j(n)$ of observations as well as the smaller numbers of inputs $i(m)$ and outputs $r(s)$ which are likely to be of interest at least in economics applications.

We shall do this in a way that should provide further conceptual clarity (and flexibility) and also facilitate our making contact with related developments in economics. First consider the following model which is the reciprocal (inefficiency) measure version of (1):

$$\min f_0 = \frac{\sum_{i=1}^{m} v_i x_{i0}}{\sum_{r=1}^{s} u_r y_{r0}} \tag{2}$$

subject to:

$$\frac{\sum_{i=1}^{m} v_i x_{ij}}{\sum_{r=1}^{s} u_r y_{rj}} \geqslant 1 ; \quad j = 1, ..., n ,$$

$$v_i, u_r \geqslant 0 .$$

Now we propose to replace these nonconvex nonlinear formulations with an *ordinary* linear programming problem. We therefore first consider

$$\max z_0 \tag{3}$$

subject to:

$$-\sum_{j=1}^{n} y_{rj}\lambda_j + y_{r0}z_0 \leqslant 0 ; \quad r = 1,..., s ,$$

$$\sum_{j=1}^{n} x_{ij}\lambda_j \leqslant x_{i0}\ ; \qquad i = 1, ..., m\ ,$$

$$\lambda_j \geqslant 0\ ; \qquad j = 1, ..., n\ .$$

Because (3) is an ordinary linear programming problem it has a linear programming dual which we can write as follows:

$$\min g_0 = \sum_{i=1}^{m} \omega_i x_{i0} \tag{4}$$

subject to:

$$-\sum_{r=1}^{s} \mu_r y_{rj} + \sum_{i=1}^{m} \omega_i x_{ij} \geqslant 0\ ,$$

$$\sum_{r=1}^{s} \mu_r y_{r0} = 1\ ,$$

$$\mu_r, \omega_i \geqslant 0\ .$$

Because of the structure of (4) one can recognize that it is equivalent to an ordinary linear fractional programming problem. (See [10] and [7].) In fact, utilizing the theory of linear fractional programming with the transformation

$$\omega_i = tv_i\ ; \quad i = 1, ..., m\ ,$$

$$\mu_r = tu_r\ ; \quad r = 1, ..., s\ ,$$

$$t^{-1} = \sum_r u_r y_{r0}\ ,$$

which, with $t > 0$, gives explicitly

$$\min f_0 = \frac{\sum_{i=1}^{s} v_i x_{i0}}{\sum_{r=1}^{s} u_r y_{r0}} \tag{6}$$

subject to:

$$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} u_r y_{rj} \geqslant 0\ ; \qquad j = 1, ..., n\ ,$$

$$v_i, u_r \geqslant 0\ ,$$

as the linear fractional programming equivalent of (4). By very evident manipulations, however, we can see that (6) is the same as (2). Hence we can use (4) to solve (6) and therefore (2) and (1) as well. Q.E.D.

We are now in an advantageous position from several standpoints. We have a completely symmetric definition of efficiency which generalizes single output ratio definitions not only in economics but in engineering and other natural sciences. We do not need to solve the nonlinear (and nonconvex) problems in which these definitions are formalized. We need only solve the ordinary linear programming problem (4) in order to obtain both the optimal f_0^* or h_0^* and the weights v_i^*, $u_r^* \geqslant 0$, since the change in variables does not alter the value of the functional.

Thus,

$$f_0^* = g_0^* = z_0^* \tag{7.1}$$

and therefore

$$h_0^* = 1/z_0^*\ . \tag{7.2}$$

Also we have the wanted relative weights. Thus nothing more is required than the solution of (4) or (3) in order to determine whether $f_0^* > 1$ or, correspondingly, whether $h_0^* < 1$, with efficiency prevailing if and only if

$$f_0^* = h_0^* = 1\ . \tag{7.3}$$

We can also effect extensions in a variety of additional (new) directions [6]. Here, however, we prefer to make contact with various developments and also sketch a few of the ideas that are elsewhere described as Data Envelopment Analysis [7]. For this purpose we introduce

$$P_j = \begin{pmatrix} Y_j \\ X_j \end{pmatrix}; \qquad j = 1, ..., n\ , \tag{8}$$

wherein the subvector Y_j contains observed output values y_{rj}, $r = 1, ..., s$ for its components and the subvector X_j contains observed input values x_{ij}, $i = 1, ..., m$.

Now consider the following vector reformulation of (3):

$$\max z_0 \tag{9}$$

[6] For instance, we could utilize the duality theory that is now associated with fractional programming (as discussed in [18] and [25] – see also [5] and [7]) as distinguished from the duality theory of ordinary linear programming or what is sometimes called duality theory (see below) in cost and production theory.

[7] This is a method for adjusting data to prescribed theoretical requirements such as optimal production surfaces, etc., *prior* to undertaking various statistical tests for purposes of public policy analysis. See [21].