ARTICLES

EFFICIENCY ANALYSIS FOR EXOGENOUSLY FIXED INPUTS AND OUTPUTS

RAJIV D. BANKER
Carnegie-Mellon University, Pittsburgh, Pennsylvania

RICHARD C. MOREY
Duke University, Durham, North Carolina

(Received February 1984; revision received March 1985; accepted November 1985)

We evaluate, by means of mathematical programming formulations, the relative technical and scale efficiencies of decision making units (DMUs) when some of the inputs or outputs are exogenously fixed and beyond the discretionary control of DMU managers. This approach further develops the work on efficiency evaluation and on estimation of efficient production frontiers known as data envelopment analysis (DEA). We also employ the model to provide efficient input and output targets for DMU managers in a way that specifically accounts for the fixed nature of some of the inputs or outputs. We illustrate the approach, using real data, for a network of fast food restaurants.

In 1978, Charnes, Cooper and Rhodes (CCR) described a mathematical programming formulation for the empirical evaluation of relative efficiency of a Decision Making Unit (DMU) on the basis of the observed quantities of inputs and outputs for a group of similar referent DMUs. They termed this approach Data Envelopment Analysis (DEA). Banker (1980) and Banker, Charnes and Cooper (1984) (BCC) provided a formal link between DEA and estimation of efficient production frontiers via constructs employed in production economics. Unlike most of the traditional econometric approaches, DEA does not require many restrictions on the underlying production technology, or an exogenous specification of the parametric form of the production correspondences. Furthermore, it focuses primarily on the technological aspects of the production correspondences and is not dependent on assumptions about or estimates of input and output prices. This feature has opened up several interesting possibilities for applications of this empirically oriented model to evaluation of efficiencies of DMUs in the regulated sector (see Banker, Conrad and Strauss 1986, and Banker 1984), in the nonprofit sector (see Farrell and Fieldhouse 1962, Lewin and Morey 1981, Lewin, Morey and Cook 1982), and in the private sector (Banker 1985, Banker and Maindiratta 1986, Banker and Datar 1987). The approach is particularly advantageous in these settings because, in general, it is difficult to obtain competitive prices for outputs and, at times, even for inputs.

The DEA approach to efficiency measurement yields managerially useful by-products. In addition to developing relative efficiency ratings for DMUs, the efficiency models provide information about the extent to which each input consumed by an inefficient DMU could be reduced without reducing any of its outputs, yet by keeping the input mix ratios (i.e., the basic technological recipe) approximately unchanged. In practical applications where it is possible to estimate input savings, DMU managers often must deal with some inputs, e.g., level of advertising, median income in service area, number of competitors, that they do not control. In such cases, information about the extent to which an exogenously fixed input variable may be reduced is not meaningful for the DMU manager. We need to further extend the CCR and BCC models in order to estimate the extent to which the controllable or discretionary inputs can be reduced by the DMU manager while keeping the exogenously fixed inputs at their current level. This is the objective that we set for the developments described in this paper.

Subject classification: 504 exogenously fixed inputs and outputs, 638 data envelopment analysis.

0030-364X/86/3404-0513 $01.25

Banker (1980) and BCC specify explicit postulates, such as convexity and monotonicity, for the underlying production possibility set employed for estimating the relative efficiency of individual DMUs. This characterization of the production possibility set will be employed to estimate the relative efficiencies when some of the inputs are not controllable. For this purpose, we shall consider a subset of the production possibility set that may be obtained by fixing the nondiscretionary inputs at their current values. We can then examine the efficiency evaluation problem in the context of the constrained production possibility set, and thus estimate possible reductions in the consumption of the discretionary inputs, keeping exogenously fixed inputs constant.

1. A Two-Dimensional Model

To describe the special considerations involved in dealing with exogenously fixed inputs, we shall employ the type of two-dimensional representation used by CCR (p. 436) to illustrate the relationship between CCR's efficiency measure and Farrell's (1957) efficiency measure. We shall consider a situation involving a single output, a discretionary input controllable by the DMU manager, and an exogenously fixed input; for ease of presentation we assume that it is desirable to conserve all inputs and increase all outputs.

We let y_j, x_{dj} and x_{fj} for $j = 1, \ldots, n$ denote the observed values of the single output, the discretionary input, and the fixed input, respectively, for each of the N DMUs. To represent the observed data in a two-dimensional diagram, we consider a section of the production possibility set by the plane given by $y = y_A$, corresponding to the level of the output for a particular DMU, referred to as DMU-A in Figure 1. Earlier developments described by BCC enable us to express the (input) technical efficiency measure for DMU-A in terms of the following linear programming formulation:

$$\text{minimize } \{e_A - \epsilon(s_{dA}^+ + s_{fA}^+ + s_A^-)\} \qquad (1)$$

$$\text{subject to } \sum_{j=1}^{N} \gamma_j x_{ij} + s_{iA}^+ = e_A x_{iA} \quad i = d, f \qquad (2)$$

$$\sum_{j=1}^{N} \gamma_j y_j - s_A^- = y_A \qquad (3)$$

$$\sum_{j=1}^{N} \gamma_j = 1 \qquad (4)$$

$$e_A, \gamma_j, s_{dA}^+, s_{fA}^+, s_A^- \geq 0.$$

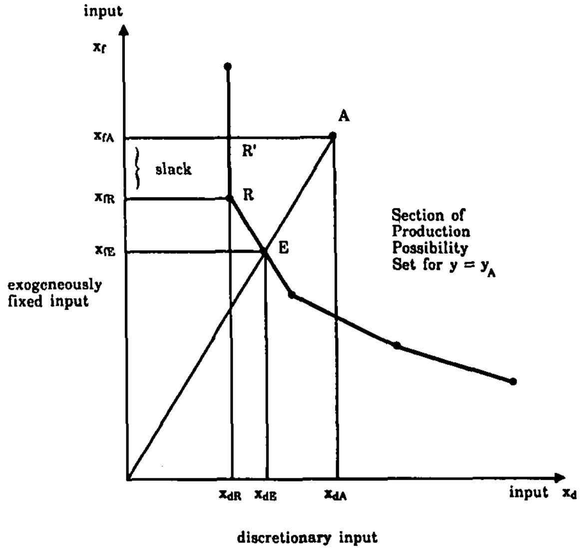

Figure 1. Maximum savings in input $x_d = |x_{dA} - x_{dR}|$. Unsubstitutable slack in input $x_f = |x_{fA} - x_{fR}| = s_{fA}^{+*}$.

In this formulation, ϵ is a small positive number. (In practice ϵ is taken to be 10^{-6}. Any theoretical difficulties associated with the use of ϵ can be avoided by formulating (1)–(4) as a multistage, preemptive series of linear programs (see, for example, Banker and Morey, 1986), where the first stage linear program minimizes e_A and the second stage goal is to maximize $(s_{dA}^+ + s_{fA}^+ + s_A^-)$.)

The reduction ξ_{iA}, possible in the consumption of each of the two inputs without reducing the level of the output (i.e., y_A), can also be estimated for DMU-A as

$$\xi_{iA} = (1 - e_A^*)x_{iA} + s_{iA}^{+*}, \quad i = d, f. \qquad (5)$$

However, since input x_f is exogenously fixed, DMU-A manager cannot reduce its consumption by the amount indicated in (5). The problem then is to determine whether and to what extent the consumption of the discretionary input x_d can be reduced, given that the level of the other input will remain fixed at its current value of x_{fA}.

To illustrate our approach, we shall refer to Figure 1. The BCC formulation evaluates the efficiency of A by comparing it with the point E on the efficient surface, and also on the contraction/expansion path of A. The coordinates of the point E are given by $x_{dE} = e_A^* x_{dA}$ and $x_{fE} = e_A^* x_{fA}$. The amount of reduction possible in the input x_d is given by the distance $|x_{dA} - x_{dE}|$. But since input x_f is exogenously fixed, information about the reduction possible in x_f is not useful to the DMU-A manager. It is therefore more meaningful to compare A with the point R on the efficient surface, which consumes the same quantity of the fixed input x_f but a smaller quantity of the discretionary input x_d. In fact, the amount of reduction in the consumption of the input x_d can be estimated as the distance $|x_{dA} - x_{dR}|$.

We shall next incorporate this restricted comparison as part of the relative efficiency evaluation procedure. Note that the terms $e_A^* x_{dA}$ and $e_A^* x_{fA}$ in (2) represent the coordinates of the point E. We shall replace these terms by the coordinates of the point R and modify (1)–(4) by considering the linear program of (6), (7), (8), (3), (4), and (9), namely:

$$\text{minimize } \{\rho_A - \epsilon(s_{dA}^+ + s_A^-)\} \quad (6)$$

$$\text{subject to } \sum_{j=1}^{N} \gamma_j x_{dj} + s_{dA}^+ = \rho_A x_{dA} \quad (7)$$

$$\sum_{j=1}^{N} \gamma_j x_{fj} + s_{fA}^+ = x_{fA} \quad (8)$$

$$\rho_A, \gamma_j, s_{dA}^+, s_{fA}^+, s_A^- \geqslant 0. \quad (9)$$

Note that in (8) we require only that the composite reference group utilize no more of the uncontrollable resource than the unit under evaluation, but not necessarily exactly the same amount; this choice enriches the comparison set. However, the prioritization ensured by the (non-Archimedean) weight ϵ, or equivalently by a multistage, preemptive series of linear programs, will first identify the maximum reduction possible in the discretionary input, before identifying the slack in the fixed input. We can then use the value ρ_A^*, and s_{dA}^{+*} obtained from the optimal solution to (3), (4) and (6)–(9) to estimate the amount of reduction possible in the discretionary input as:

$$\epsilon_{dA} = (1 - \rho_A^*) x_{dA} + s_{dA}^{+*}. \quad (10)$$

Furthermore, even after we affect this reduction in the discretionary input there may be slack s_{fA}^+ for the fixed input. Since the maximum reduction in the discretionary input, namely ρ_A, has already been evaluated via the minimization problem in (6), the slack represents the quantity of the fixed input that cannot be substituted for any further reduction of the discretionary input. (Note that the objective function in (6) included slacks only for the discretionary input and the output so as to insure that no priority is given to any slack associated with nondiscretionary input.) This property is reflected in the dual only as nonnegativity requirements for the associated dual variable, rather than a strict positivity requirement.

As noted in CCR and BCC, because of its special structure, the dual for the previous problem can be expressed as a fractional programming problem, i.e.,

$$\text{maximize } \frac{u y_A - v_f x_{fA} - u_0}{v_d x_{dA}} \quad (11)$$

$$\text{subject to } \frac{u y_j - v_f x_{fj} - u_0}{v_d x_{dj}} \leqslant 1, \quad j = 1, \ldots, n \quad (12)$$

$$u, v_d \geqslant \epsilon > 0, \quad v_f \geqslant 0, \quad (13)$$

and u_0 is unconstrained in sign (it is the dual variable associated with (4)). Note that (11) involves the ratio of weighted outputs to weighted discretionary inputs and hence is similar to the original CCR formulation (p. 430); this reformulation completes our desired extension of the relative technical efficiency measure for the two-dimensional case.

2. Generalization to Multiple Outputs and Multiple Inputs

In this section we develop a general model for situations involving multiple outputs, multiple discretionary inputs and multiple exogenously fixed inputs. As in the preceding section, we shall also estimate the amount of reduction possible in the consumption of discretionary inputs by interpolating the performance of the most efficient DMUs, employing, as in BCC, the characterization of the production possibility set in terms of convexity, monotonicity and other properties. This approach identifies any relative technical inefficiencies that may be present at the level of operations of a DMU_0 that is being evaluated.

We let y_{rj} for $r \in \{1, \ldots, s\}$ and x_{ij} for $i \in I = \{1, \ldots, M\}$ be the observed outputs and inputs values for $j = 1, \ldots, N$ DMUs. We shall categorize each input as either discretionary, and assign it to an index set I_D, or fixed, and assign it to an index set I_F, so that $I_D \cup I_f = I = \{1, \ldots, M\}$ and $I_D \cap I_f = \varnothing$.

The BCC formulation to measure the technical (input) efficiency measure η for a DMU_0 is given by the following linear programming problem:

$$\eta_0^T = \min \left\{ e_0 - \epsilon \left(\sum_{i=1}^{M} s_{io}^+ + \sum_{r=1}^{S} s_{ro}^- \right) \right\} \quad (14)$$

subject to

$$\sum_{j=1}^{N} \gamma_j x_{ij} + s_{i0}^+ = e_0 x_{i0} \quad i \in \{1, \ldots M\} \quad (15)$$

$$\sum_{j=1}^{N} \gamma_j y_{rj} - s_{r0}^- = y_{r0} \quad r \in \{1, \ldots s\} \quad (16)$$

$$\sum_{j=1}^{N} \gamma_j = 1 \tag{17}$$

$$e_0,\ \gamma_j,\ s_{i0}^+,\ s_{r0}^- \geq 0. \tag{18}$$

The quantity $\epsilon > 0$ is as defined in (1)–(4).

Proceeding as in Section 1, we now modify the constraints corresponding to the exogenously fixed inputs to reflect the fact that these inputs cannot be reduced, so that we can estimate the maximum reduction possible in the other inputs. We therefore present the modified technical evaluation problem as follows:

$$\eta_0^{MT} = \min \left[\rho_0 - \epsilon\left(\sum_{i\in I_D} s_{i0}^+ + \sum_{r=1}^{S} s_{r0}^-\right)\right] \tag{19}$$

subject to

$$\sum_{j=1}^{N} \gamma_j x_{ij} + s_{i0}^+ = \rho_0 x_{i0} \quad i \in I_D \tag{20}$$

$$\sum_{j=1}^{N} \gamma_j x_{ij} + s_{i0}^+ = x_{i0} \quad i \in I_F \tag{21}$$

$$\sum_{j=1}^{N} \gamma_j y_{rj} - s_{r0}^- = y_{r0} \quad r \in \{1, \ldots, s\} \tag{22}$$

$$\sum_{j=1}^{N} \gamma_j = 1 \tag{23}$$

$$\rho_0,\ \gamma_j,\ s_{i0}^+,\ s_{i0}^- \geq 0. \tag{24}$$

The dual of this linear programming problem has a form that enables us to represent it as a fractional programming problem.

$$\eta_0^{MT} = \max \frac{\sum_{r=1}^{S} u_r y_{r0} - \sum_{i\in I_F} v_i x_{i0} - u_0}{\sum_{i\in I_D} v_i x_{i0}} \tag{25}$$

subject to

$$\frac{\sum_{r=1}^{S} u_r y_{rj} - \sum_{i\in I_F} v_i x_{ij} - u_0}{\sum_{i\in I_D} v_i x_{ij}} \leq 1,$$

$$j \in \{1, \ldots N\} \tag{26}$$

$$u_r, v_i \geq \epsilon > 0, \quad \text{for} \quad r \in \{1, \ldots s\} \quad \text{and} \quad i \in I_D \tag{27}$$

$$v_i \geq 0 \quad \text{for} \quad i \in I_F \tag{28}$$

and u_0 is unconstrained in sign. (29)

The modified technical efficiency measure from (19)–(24) is directly related to the BCC technical efficiency measure from (14)–(18). From the diagram in Figure 1, it is visually apparent that

$$e_0^* = \frac{OE}{OA} = \frac{x_{dE}}{x_{dA}} \quad \text{is greater than} \quad \rho_0^* = \frac{x_{dR}}{x_{dA}},$$

at least for the one output, one discretionary input and one fixed input case depicted. A similar relationship between these measures is valid even for the general case discussed in this section. We formalize this relation in terms of the following result:

Proposition 1. *The modified technical efficiency component ρ_0^* from* (19)–(24) *for any DMU_0 is less than or equal to the corresponding technical efficiency component e_0^* from* (14)–(18).

To prove this proposition, we note first that $e_0^* \leq 1$ in optimal solution to (14)–(18) because DMU_0 is itself one of the $j = 1, \ldots, N$ referent observations. By comparing the constraint sets in the two linear programs, we see that any optimal solution to (14)–(18) is a feasible solution for (19)–(24); hence, $\rho_0^* \leq e_0^*$.

3. Technical and Scale Efficiency

Our analysis so far has addressed only the measurement of the purely technical efficiency at the given scale of operation for DMU_0. BCC have proved that the original CCR formulation obtained by deleting the constraint $\sum_{j=1}^{N} \gamma_j = 1$ in the programming problem of (14)–(18) provides a measure for the aggregate technical and scale efficiency of DMU_0. Therefore, to estimate the modified aggregate technical and scale efficiency measure when some of the inputs are fixed, we write the following programming problem:

$$\eta_0^{MA} = \min \left\{h_0 - \epsilon\left(\sum_{i\in I_D} s_{i0}^+ + \sum_{r=1}^{S} s_{r0}^-\right)\right\} \tag{30}$$

$$\text{subject to} \quad \sum_{j=1}^{N} \mu_j x_{ij} + s_{i0}^+ = h_0 x_{i0} \quad i \in I_D \tag{31}$$

$$\sum_{j=1}^{N} \mu_j x_{ij} + s_{i0}^+ = \sum_{j=1}^{N} \mu_j x_{i0} \quad i \in I_F \tag{32}$$

$$\sum_{j=1}^{N} \mu_j y_{rj} - s_{r0}^- = y_{r0} \quad r \in \{1, \ldots s\} \tag{33}$$

$$h_0, \mu_j, s_{i0}^+, s_{r0}^- \geq 0, \tag{34}$$

and $\epsilon > 0$, as in (1)–(4).

Hence, if the level of the ith discretionary input for the DMU under evaluation were adjusted to the level

$$x'_{i0} = \frac{h_0^* x_{i0} - s_{i0}^{+*}}{\sum_{j=1}^{N} \mu_j^*} \quad i \in I_D, \tag{35}$$

and the level of the rth output for the DMU under evaluation were adjusted to the level of

$$y'_{r0} = \frac{y_{r0} + s_{r0}^{-*}}{\sum_{j=1}^{N} \mu_j^*} \quad r \in \{1, 2, \ldots, s\}, \tag{36}$$

then the DMU would be technically efficient and operate at the most productive scale size so as to have a modified aggregate efficiency rating of one.

The dual of this linear program can also be cast in the form of a linear fractional program similar to the original CCR formulation:

$$\eta_0^{MA} = \max \frac{\sum_{r=1}^{S} u_r y_{r0}}{\sum_{i \in I_D} v_i x_{i0}} \tag{37}$$

subject to

$$\frac{\sum_{r=1}^{S} u_r y_{rj} - \sum_{i \in I_F} v_i (x_{ij} - x_{i0})}{\sum_{i \in I_D} v_i x_{ij}} \leq 1 \quad j \in \{1, \ldots N\} \tag{38}$$

$$u_r, v_i \geq \epsilon > 0 \quad \text{for} \quad r \in \{1, \ldots s\} \quad \text{and} \quad i \in I_D \tag{39}$$

$$v_i \geq 0 \quad \text{for} \quad i \in I_F. \tag{40}$$

As in BCC, we now present a formal relationship between the modified technical efficiency measure in (19)–(24) and the modified aggregate technical and scale efficiency measure from (30)–(34).

Proposition 2. *The modified technical efficiency measure* η_0^{MT} *is greater than or equal to the modified aggregate technical and scale efficiency measure* η_0^{MA}.

The result follows immediately from the observation that an optimal solution for the former linear program is feasible for the latter linear program.

The modified scale efficiency measured as the ratio of the modified aggregate efficiency to the modified technical efficiency is, therefore, less than or equal to one, and it reflects the divergence from the most productive scale size for the controllable inputs. In other words, if $\sum_{j=1}^{N} \mu_j^* = 1$ for an optimal solution to (30)–(34) then the modified scale efficiency rating for the DMU_0 is equal to one, and it implies that DMU_0 is currently operating at the most productive scale size for the discretionary inputs, given the fixed level of the non-discretionary inputs. Further, as in Banker (1984), $\sum_{j=1}^{N} \mu_j^* > 1$ implies that DMU_0 is operating at a scale greater than the most productive scale size for the discretionary inputs. It is also possible to infer, therefore, that the *technically* efficient referent point given by $y_r' = y_{r0} + s_{r0}^{-*}$, $x_i' = x_{i0} - s_{i0}^{+*}$ for $i \in I_F$, and $x_i' = \rho_0^* x_{i0} - s_{i0}^{+*}$ for $i \in I_D$, is in the region of the production possibility set where decreasing returns to scale prevail. Conversely, if $\sum_{j=1}^{N} \mu_j^* < 1$, then DMU_0 is operating in the increasing returns to scale region, at a scale smaller than the most productive scale size for the discretionary inputs, given the fixed level of the non-discretionary inputs. The managerial implication of this inference is that average productivity can potentially be increased by increasing the level of activity in terms of the outputs y_r by *proportionately increasing* consumption of the discretionary inputs x_i for $i \in I_D$ as indicated in (35), while keeping the non-discretionary inputs x_i for $i \in I_F$ fixed at their current levels.

4. Output Targets

So far we have focused on the problem of estimating the extent to which we could reduce the consumption of discretionary inputs without reducing output production. An alternative approach could consider the estimation of the extent to which output production could be increased without requiring additional inputs. For this purpose, we shall consider the following programming problem to estimate the (output) technical inefficiency measure for a DMU_0.

$$\text{Maximize} \left\{ g_0 + \epsilon \left(\sum_{r=1}^{S} s_{r0}^- + \sum_{r=1}^{M} s_{i0}^+ \right) \right\} \tag{41}$$

subject to

$$\sum_{j=1}^{N} \gamma_j y_{rj} - s_{r0}^- = g_0 y_{r0}, \quad r \in \{1, \ldots, s\} \tag{42}$$

$$\sum_{j=1}^{N} \gamma_j x_{ij} + s_{i0}^+ = x_{i0}, \quad i \in \{1, \ldots, M\} \tag{43}$$

$$\sum_{j=1}^{N} \gamma_j = 1 \tag{44}$$

$$g_0, \gamma_j, s_{r0}^-, s_{i0}^+ \geq 0. \tag{45}$$

We note that the M constraints corresponding to the M inputs compare the reference input quantities $\sum_{j=1}^{N} \gamma_j x_{ij}$ with the corresponding quantities x_{i0} for DMU_0, and *not* proportionately reduced quantities $e_0 x_{i0}$ as in (14)–(18). Therefore, these targets project a reduction in the inputs of only the slack quantities s_{i0}^+. By modifying the objective function of the programming problem in (41)–(45) to read

$$\text{maximize} \left\{ g_0 + \epsilon \left(\sum_{r=1}^{S} s_{r0}^- + \sum_{i \in I_D} s_{i0}^+ \right) + \epsilon' \sum_{i \in I_F} s_{i0}^+ \right\}, \tag{46}$$

where ϵ and ϵ'/ϵ are small (non-Archimedean) positive numbers, we can substitute as much of the factors s_{i0}^+ for $i \in I_F$ as possible, in order to secure further reductions in the targets for the discretionary inputs, or increases in the targets for the outputs. Any positive s_{i0}^+ for $i \in I_F$ then represent surplus quantities of fixed inputs that cannot be substituted for a discretionary input or used to produce more output in a manner consistent with the underlying production technology reflected in the observed input and output level.

5. Exogenously Fixed Outputs

The analysis described in the preceding sections can be directly extended to situations with some exogenously fixed multiple outputs that are therefore beyond the discretionary control of the DMU manager. An example of an exogenously determined output may include check cashing transactions in a bank, where output is related to the performance of a purely gratis service function. In this case, management is interested in determining the maximum factor by which only the *controllable* outputs can be augmented without requiring any more resources. We shall, once again, consider the programming formulation (41)–(45). If all the outputs were under the discretionary control of the DMU manager, then the adjustments needed in all of the outputs to render the DMU technically efficient is given by

$$y'_{r0} = (g_0^* y_{r0} + s_{r0}^{-*}) \quad \text{for} \quad r = 1, 2, \ldots, S. \tag{47}$$

However, in situations with the levels of some outputs exogenously fixed, beyond the discretionary control of the DMU manager, these targets for increasing outputs may not be meaningful. Proceeding in a manner analogous to our earlier analysis of exogenously fixed inputs, we seek to estimate the maximum possible increase in the discretionary outputs, keeping the inputs and exogenously fixed outputs at their current levels. Therefore, we write the following programming problem:

$$\text{maximize} \left\{ \varnothing_0 + \epsilon \left(\sum_{r \in \nu_D} s_{r0}^- + \sum_{i=1}^{M} s_{i0}^+ \right) \right\} \tag{48}$$

subject to

$$\sum_{j=1}^{N} \gamma_j y_{rj} - s_{r0}^- = \varnothing_0 y_{r0} \quad r \in \nu_D \tag{49}$$

$$\sum_{j=1}^{N} \gamma_j y_{rj} - s_{r0}^- = y_{r0} \quad r \in \nu_F \tag{50}$$

$$\sum_{j=1}^{N} \gamma_j x_{ij} + s_{i0}^+ = x_{i0} \quad i \in \{1, \ldots, M\} \tag{51}$$

$$\sum_{j=1}^{N} \gamma_j = 1 \tag{52}$$

$$\varnothing_0, \gamma_j, s_{r0}^-, s_{i0}^+ \geq 0. \tag{53}$$

In this formulation ν_D is the set of indices of outputs under the discretionary control of the DMU manager and ν_F is the set of indices of exogenously fixed outputs. The maximum possible increases in the discretionary outputs are then given by (47).

Further, the dual of the programming problem in (48)–(53) can be cast in the form of the following fractional programming problem:

$$\text{minimize} \ \frac{\sum_{i=1}^{M} v_i x_{i0} - \sum_{r \in \nu_F} u_r y_{r0} - u_0}{\sum_{r \in \nu_D} u_r y_{rj}}$$

subject to

$$\frac{\sum_{i=1}^{M} v_i x_{ij} - \sum_{r \in \nu_F} u_r y_{rj} - u_0}{\sum_{r \in \nu_D} u_r y_{rj}} \leq 1, \quad j = 1, \ldots, N,$$

$v_i \geq \epsilon$ for $i \in \{1, 2, \ldots, M\}$, $u_r \geq \epsilon$ for $r \in \nu_D$, $u_r \geq 0$ for $r \in \nu_F$. This problem is similar to the original formulation for estimating the relative (output) inefficiency of a DMU. As before, the relationship between this modified technical inefficiency measure and the original technical inefficiency measure can be formalized by means of the following proposition:

Proposition 3. *The modified (output) technical inefficiency component $\varnothing_0$ obtained from (48)–(53) for any DMU$_0$ is greater than or equal to the corresponding (output) technical inefficiency component g_0 obtained from (41)–(45).*

The proof of this proposition is analogous to the proof of Proposition 1 and is omitted.

6. Empirical Illustration

In this section we shall apply the analytical techniques developed in this paper to a real data set for a 60-DMU network of fast food restaurants. This example will also serve to illustrate the impact of fixed uncontrollable inputs; in particular, it will compare the difference in controllable input targets that results from two different treatments of the non-discretionary inputs—one in which all inputs are indeed treated as discretionary, and the other in which only the truly controllable inputs are treated as discretionary. It will show the differences in targets that result from eliminating only technical inefficiences and from eliminating both technical and scale inefficiences.

We considered data on 3 outputs and 6 inputs for 60 restaurants in the fast food chain. The 3 outputs, all considered controllable, corresponded to the 3 principal categories of food sales for the fast food chain: breakfast, lunch and dinner sales. Two of the 6 inputs corresponded to expenditures for supplies and materials, and expenditures related to labor; these were clearly discretionary. Two other input variables represented the age of the store and the advertising expenditures allocated to the store by headquarters. Decisions regarding advertising expenditures were made at the corporate level, and therefore, this input

Table I
Comparison of Various Inefficiency Measures for DMU-51

Measure	All 6 Inputs Treated as Discretionary	Only Inputs 1 and 2 Treated as Discretionary
1. Technical efficiency	0.964	0.904
2. Aggregate technical and scale efficiency	0.948	0.899
3. Scale efficiency	0.982	0.994
4. Returns to scale	Increasing returns: $\sum_{j=1}^{60} \mu_j^* = .9399$	Decreasing returns: $\sum_{j=1}^{60} \mu_j^* = 1.0542$

was beyond the control of the individual restaurant manager. The final 2 inputs were demographic and indicated whether the store was located in an urban or rural area, and whether it had a drive-in window. These uncontrollable characterizations were handled by means of binary 0–1 variables, with zeroes corresponding, respectively, to a rural location and a location with no drive-in window. For evaluating a store located in a rural area or one without a drive-in, these variables force the peer group members (i.e., members of the reference group) to be compared only to like stores. However, the reference group for stores in urban areas (considered to be a more favorable environment), or stores with drive-ins, could consist of stores of both types. (Banker and Morey 1986 treat the use of categorical variables to model situations lacking a continuous scale. They treat both controllable and uncontrollable categorical variables, the controllable categorical variables necessitating a mixed-integer LP formulation.) The detailed data set for the 60 stores is available from the authors.

In Table I, we compare in detail the results of the efficiency analyses, performed under different treatments, for one of the stores, DMU-51. Treatment I considered all 6 inputs to be controllable (the only modeling possibility previously available, i.e., formulation (14)–(18)). The result was an optimal proportionate reduction factor e_0^* of 0.965, which implies a 3.5% simultaneous reduction target for all 6 inputs. In contrast, treatment II considers only the first two inputs to be controllable; in this case the maximum possible proportionate reduction factor ρ_0^*, obtained from formulation (19)–(24), is 0.904. This result implies a 9.6% reduction target for these two controllable inputs, without requiring the remaining four inputs to be lowered from their current levels. Note that this outcome is consistent with Proposition 1.

It is also interesting to compare for the 51st DMU the estimates of returns to scale for each of the 2 treatments. In both cases, if we utilize (30)–(34), divergence from the most productive scale size is indicated. However, when we consider all the inputs to be discretionary, $\sum_{j=1}^{60} \mu_j^* = 1.0542$ indicates that DMU-51 is operating in an *increasing* returns-to-scale segment of the production possibilities set. In other words, if we increased *all* 6 inputs proportionately, then we could increase 3 outputs in a greater proportion. However, when only the first 2 inputs were considered discretionary, *decreasing* returns to scale were indicated because $\sum_{j=1}^{60} \mu_j^* = 0.9399$ is less than one. This result implies that if only the first two inputs were increased proportionately, keeping the four nondiscretionary inputs fixed at their current levels, then the three outputs would be increased by a small proportion.

These considerations are reflected in the targets displayed in Table II for DMU-51 to eliminate technical, or technical and scale, inefficiencies. First, we observe that the targets suggested (from (10)) for inputs 1 and 2 for eliminating technical inefficiency

Table II
Targets for DMU-51 to Eliminate Inefficiencies

	Current Level	Target for Eliminating Technical Inefficiencies		Targets for Eliminating Both Technical and Scale Inefficiencies	
		All inputs considered discretionary	Only inputs 1 and 2 considered discretionary	All inputs considered discretionary	Only inputs 1 and 2 considered discretionary
Input 1, cost of supplies and goods	\$254.39K	\$245.91K	\$230.08K	\$256.63K	\$216.99K
Input 2, cost of labor	\$205.06K	\$181.18K	\$179.80K	\$185.88K	\$174.91K
Output 1	\$165.75K	\$165.25K	\$165.25K	\$175.82K	\$156.75K
Output 2	\$489.77K	\$489.77K	\$489.77K	\$521.09K	\$465.59K
Output 3	\$ 99.57K	\$149.39K	\$135.26K	\$163.89K	\$125.01K

Table III
Impacts of Different Assumptions About Controllable Inputs on Inefficiency Measurement

Measure	All 6 Inputs Considered Controllable	Only 2 of the 6 Inputs Considered Controllable
1. No. of units scored technically efficient	33	33
2. Range of technical inefficiency scores	0.8769–1.0	0.8725–1.0
3. No. of units scored technically and scale efficient	24	32
4. Range of technical and scale inefficiency scores	0.7665–1.0	0.8766–1.0

(columns 2 and 3 of Table II) are more demanding when only these 2 inputs are considered discretionary. Next, we consider the elimination of both technical and scale inefficiencies (columns 4 and 5 of Table II, using formula (35)). When all inputs are considered discretionary, the reduction required in the inputs to eliminate technical inefficiency is offset by increases required to exploit the prevailing increasing returns to scale. Therefore, the targets were higher than those required to eliminate only technical inefficiency. (The net target associated with the cost of labor of $185.88K (column 4, row 2) is a reduction (in spite of the increasing returns to scale), since from (35), the scaling occurs after the inefficiencies have been eliminated.) On the other hand, when only inputs 1 and 2 were considered discretionary, reductions were required in the two inputs to eliminate technical inefficiency as well as scale inefficiency, because a decreasing return to scale prevailed in this case. Consequently, the targets for the two inputs were lower than those required to eliminate only technical inefficiency.

We also observed similar impacts on inefficiency measurement for the other 35 DMUs that were *not* rated technical and scale efficient (see Table III). There was no change in the efficiency classification of the 24 DMUs that were operating on the efficient production frontier and at the most productive scale size. But the impact on the inefficiency score increased for the more inefficient DMUs, as is evident from the data presented in Tables IV and V. The mean absolute change in the technical and scale inefficiency score was 7.98% for the 11 DMUs with inefficiency rating below 0.900.

Table IV
Mean Absolute Change in Technical and Scale Inefficiency Scores when Only 2 Inputs Are Considered Controllable Instead of All 6 Inputs

Range of Inefficiency Score with All 6 Inputs Considered Controllable	No. of Units in Each Category	Mean Absolute Value of Change in Inefficiency Score
1. 1.0	24	0
2. 0.950–0.999	10	0.0168
3. 0.900–0.949	15	0.0298
4. <0.900	11	0.0795

Table V
Mean Absolute Change in Targets to Attain Technical and Scale Efficiency when Only 2 Inputs Instead of All 6 Are Considered Controllable

	Input or Output	Mean Absolute Change in Target	Mean Absolute Percentage	Largest Change in Target (over all 36 units)
I.1	Cost of supplies and materials	$20,067	8.09%	$120,000
I.2	Cost of labor	9,882	5.34%	63,200
O.1	Sales of type 1	10,640	8.94%	71,000
O.2	Sales of type 2	53,460	9.81%	402,000
O.3	Sales of type 3	23,403	15.29%	167,000

The impact on the targets for the 36 inefficient DMUs, if they were to attain technical and scale efficiency, was also considerable. The mean absolute percentage change for the 36 DMUs ranged from 5.36% to 15.29% for the output "sales of type 3." It is thus apparent from the empirical application that reflecting the fixed nature of some of the inputs permits the identification of a considerably enhanced opportunity for targeted savings in the controllable inputs or for targeted increases in the controllable outputs.

Acknowledgment

Parts of this paper rest on developments in Rajiv D. Banker's doctoral thesis, "Studies in Cost Allocation and Efficiency Evaluation," at the Graduate School of Business Administration, Harvard University. Helpful comments by W. W. Cooper are gratefully acknowledged.

References

BANKER, R. D. 1980. Studies in Cost Allocation and Efficiency Evaluation. Unpublished doctoral thesis, Harvard University, Graduate School of Business Administration.

BANKER, R. D. 1984. Estimating Most Productive Scale Size Using Data Envelopment Analysis. *Eur. J. Opnl. Res.* 35–44.

BANKER, R. D. 1985. Productivity Measurement and Management Control. In *The Management of Productivity and Technology in Manufacturing*, P. Kleindorfer (ed.). Plenum Publishing, New York.

BANKER, R. D., AND S. DATAR. 1987. Accounting for Manufacturing Operations and Labor Productivity: An Application. In *Field Studies in Management Accounting and Control*, W. Bruns and R. Kaplan (eds.). Harvard Business School Press, Cambridge Mass.

BANKER, R. D., AND A. MAINDIRATTA. 1986. Piecewise Loglinear Estimation of Efficient Production Surfaces. *Mgmt. Sci.* **32,** 123–135.

BANKER, R. D., AND R. C. MOREY. 1986. The Use of Categorical Variables in Data Envelopment Analysis. *Management Science* **35,** December.

BANKER, R. D., A. CHARNES AND W. W. COOPER. 1984. Models for Estimation of Technical and Scale Inefficiencies in Data Envelopment Analysis. *Mgmt. Sci.* **30,** 1078–1092.

BANKER, R. D., R. F. CONRAD AND R. P. STRAUSS. An Application of Data Envelopment Analysis to the Empirical Investigation of a Hospital Production Function. *Mgmt. Sci.* **32,** 30–44.

CHARNES, A., W. W. COOPER AND E. RHODES. 1978. Measuring the Efficiency of Decision Making Units. *Eur. J. Opnl. Res.* 429–444.

FARRELL, M. J. 1957. The Measurement of Productive Efficiency. *J. Roy. Statist. Soc.* Ser. A, pp. 253–290.

FARRELL, M. J., AND M. FIELDHOUSE. 1962. Estimating Efficient Production Frontiers Under Increasing Returns to Scale. *J. Roy. Statist. Soc.* Ser. A, pp. 252–67.

LEWIN, A. Y., AND R. C. MOREY. 1981. Measuring the Relative Efficiency of Output Potential of Public Sector Organizations: An Application of Data Envelopment Analysis. *J. Policy Anal. Inform. Syst.* (December).

LEWIN, A. Y., R. C. MOREY AND T. J. COOK. 1982. Evaluating the Administrative Efficiency of Courts. *Omega.*