Journal of Productivity Analysis, 9, 205–232 (1998)

Marginal Rates and Two-dimensional Level Curves in DEA

DAN ROSEN drosen@algorithmics.com
Centre for Management of Technology and Entrepreneurship, Dept. of Industrial Engineering, University of Toronto, 4 Taddle Creek Rd., Toronto, Ontario, M5S 1A4, CANADA

CLAIRE SCHAFFNIT schaffn@ie.utoronto.ca
Centre for Management of Technology and Entrepreneurship, Dept. of Industrial Engineering, University of Toronto, 4 Taddle Creek Rd., Toronto, Ontario, M5S 1A4, CANADA

JOSEPH C. PARADI paradi@ie.utoronto.ca
Centre for Management of Technology and Entrepreneurship, Dept. of Industrial Engineering, University of Toronto, 4 Taddle Creek Rd., Toronto, Ontario, M5S 1A4, CANADA

Abstract

Of great importance to management, the computation of *trade-offs* presents particular difficulties within DEA since the piecewise linear nature of the envelopment surfaces does not allow for unique derivatives at every point. We present a comprehensive framework for analyzing marginal rates, and directional derivatives in general, on DEA frontiers. A useful characterization of these derivatives at given points can be provided in terms of the ranges they can take; equivalently, the bounds of these ranges correspond to derivatives "to the right" and "to the left" at these points. We present two approaches for their computation: first, the dual equivalents calculation of minimum and maximum multiplier ratios / finite differences, and then a modified simplex tableau method. The simplex tableau method provides a more general application of the method introduced by Hackman et al. (1994) to generate *any* two-dimensional section of the isoquant and is a practical tool to generate level plots of the frontier. By giving a complete picture of trade-offs and allowing a better visualization of high dimensional production possibility sets, these tools can be very useful for managerial applications.

Keywords: Data Envelopment Analysis, marginal rates, trade-offs, partial derivatives, piecewise linear surfaces, returns to scale

1. Introduction

Of major importance to management is the knowledge of *trade-offs* implicit in a production process. For instance, managers often want to know the additional amount of a certain input that is required in order to increase a particular output by a (small) fixed amount. An example of this was given in Thanassoulis (1993), where the marginal costs of outputs (teaching units, regular patients and severe patients) for a sample of hospitals were estimated using Data Envelopment Analysis (DEA) and regression analysis. Similarly, managers are often interested in the required increase in one input when another input is decreased; or in the reduction of one output that is necessary in order to increase another output, while the rest of the factors are kept constant. For example, in the case of software production, it is important to estimate the decrease in software quality (*e.g*. increase in bug count or in rework hours) if the program size increases and everything else (cost, time to market) stays constant; *c.f*. Paradi et al. (1995).

The computation of trade-offs presents special difficulties within DEA. Mathematically, these trade-offs (marginal rates of substitution, marginal rates of transformation, marginal costs, marginal productivities) represent partial derivatives on the frontier; *i.e*. slopes on the production frontier at particular points. DEA constructs a piecewise linear surface as an approximation of the frontier; hence, the DEA surface is continuous but may have discontinuous derivatives (*e.g*. at the extreme efficient DMUs). As has been widely recognized, this translates to multiple optimal solutions for the DEA multipliers. Although the ratios of pairs of multipliers give the desired slopes on a particular facet of the frontier (see Charnes et al., 1985), their non-uniqueness causes problems of interpretation in practice. In fact, the multipliers that result from the LP solutions have been incorrectly used in some publications, without further noting the arbitrariness of such a decision. Other authors have pointed to the multiplicity problem but, without another practical alternative, still used the multipliers from the LP solutions in their analyses (*c.f*. Thanassoulis, 1993; Chilingerian, 1995).

It is important to emphasize that *the discontinuity of the slopes on the envelopment surface is an intrinsic property of the DEA methodology*. Hence, a complete characterization of trade-offs would require the identification of all the frontier facets and their intersections. With the recent introduction of efficient algorithms and software to obtain all the facets of closed convex cones (*c.f*. Fukuda, 1993), such an approach seems appealing and certainly practical for small problems. However, as the number of facets grows exponentially with the dimensionality of the problem, so does the number values for each trade-off at any given extreme efficient point. The picture becomes gloomier when we consider that the number of trade-offs one can investigate also increases drastically. Moreover, such an approach does not directly discern whether movement in a particular direction along a facet is feasible from a given point on that facet. Then, it seems that a simpler and more productive approach is to characterize the slopes at certain points neither as single nor multiple numbers but, more generally, in terms of their upper and lower bounds. Equivalently, at a given point, the (non-unique) partial derivatives may be characterized as *derivatives to the right or to the left* of the point; *i.e*. as limits from different directions.

In this paper, we present two approaches for obtaining these derivatives: (i) direct calculation of the *minimum and maximum ratios* of any two multipliers, and (ii) a *modified simplex tableau* method.

The first method requires the solution of two linear programs; hence it can be readily solved without much programming effort. We further show that, with very simple variable transformations, its dual formulation is equivalent to a *finite differences* approach.

In contrast, the simplex tableau method requires the (albeit simple) modification of existing LP codes for its solution. However, with the equivalent cost of two LP solutions, this method can generate *any* two-dimensional section of the isoquant and, thus, effectively computes the derivatives at each vertex of the section, as well. This algorithm can be used further to generate 2-D sections of the frontier at different levels of the other variables, similar to the level plots used in various areas of engineering and physical sciences. By allowing a better visualization of production possibility sets for high dimensional problems, this tool can be of great practical importance for managerial applications. As is shown in Section 5, for any two variables, this technique provides more realistic 2-D snapshots that take into account the levels of *all* the variables. In contrast, the simple plot of the *raw* 2-D data for each DMU can be seriously misleading for multidimensional problems.

It should be noted that the non-uniqueness of optimal multipliers has also caused difficulties in the assessment of returns to scale (RTS) (Banker et al., 1984). Banker and Thrall (1992) resolved this problem with an algorithm to compute the minimum and maximum value of the (unconstrained) RTS multiplier. More recently, Golany and Yu (1994) used an approach akin to finite differences to characterize the RTS "to the right" and "to the left" of an efficient point, while Hackman et al. (1994), presented an algorithm based on the simplex tableau to represent "radial" two-dimensional sections of the DEA frontier. This method corresponds to the computation of RTS intervals, at different levels of input (or output) for a certain input-output mix. In this work, we show that the algorithms presented in these three papers can be applied more generally to characterize *all* slopes on the envelopment surface, and we further present them in a unified way. In fact, the analyses of RTS and marginal rates can be seen as specific cases of the general formulation of directional derivatives on the frontier, as is shown in Section 6.

The rest of the paper is organized as follows: in sections 2 & 3, we introduce the basic ideas behind DEA and a formal presentation of the problem. The three formulations and computational algorithms are presented in section 4, and application examples are given in section 5. Section 6 presents the generalization of the approach to arbitrary directions, and Section 7 closes with some concluding remarks.

2. DEA Production Frontiers

Consider a set of n DMUs, each consuming different amounts of m inputs to produce s outputs. Let $\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{mk})^{\mathrm{T}} \in \Re^m_+$ and $\mathbf{y}_k = (y_{1k}, y_{2k}, \ldots, y_{sk})^{\mathrm{T}} \in \Re^s_+$ denote the inputs and outputs of the k-th DMU, $k = 1, 2, \ldots, n$. We assume that each DMU has at least one positive input and one positive output.

From these observations, we can construct the production possibility set

$$T = \left\{ (\mathbf{x}, \mathbf{y}) : \mathbf{x} \geq \sum_{k=1}^{n} \lambda_k \mathbf{x}_k, \mathbf{y} \leq \sum_{k=1}^{n} \lambda_k \mathbf{y}_k, \sum_{k=1}^{n} \lambda_k = 1; \lambda_k \geq 0 \, \forall \, k \right\} \tag{2.1}$$

This is the smallest convex set that satisfies free disposability of inputs and outputs and includes all the observations (more formally, it satisfies the postulates of convexity, monotonicity, inclusion and minimum extrapolation; *c.f.* Banker and Thrall (1992)).

Given this set, the efficiency of a particular DMU, call it DMU_o , can be evaluated through the following linear program (referred to as the BCC model, Banker et al. (1984))

$$\begin{array}{lll} \min_{\theta,\lambda} & \theta & \\ \text{s.t.} & \mathbf{x}_o\theta - & \mathbf{X}\lambda \geq \mathbf{0} \\ & & \mathbf{Y}\lambda \geq \mathbf{y}_o \\ & & \mathbf{e}^{\mathrm{T}}\lambda = 1;\ \ \lambda_k \geq 0\ \ \forall k,\ \theta \text{ free} \end{array} \tag{2.2}$$

where, for simplicity of notation, we denote by $\mathbf{X}$ the matrix of inputs, with columns $\mathbf{x}_k$, and by $\mathbf{Y}$ the output matrix, with columns $\mathbf{y}_k$; $\mathbf{e}$ is a vector of ones, and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)^{\mathrm{T}}$.

The optimal θ in (2.2) is an *input radial measure of technical efficiency*, and can be seen as the DEA variant of the Debreu-Farrell measure. The non-zero elements of λ identify the *reference set* of DMU_o: the “dominating” DMUs located on the frontier against which the inefficient unit is evaluated.

The dual program of (2.2) is given by

$$\begin{array}{lll} \max_{\mu,\nu,\omega} & \mathbf{y}_0^{\mathrm{T}}\mu + \omega & \\ \text{s.t.} & \mathbf{x}_0^{\mathrm{T}}\nu & = 1 \\ & \mathbf{Y}^{\mathrm{T}}\mu - \mathbf{X}^{\mathrm{T}}\nu + \omega & \leq \mathbf{0} \\ & & \mu_i, \nu_j \geq \mathbf{0}\ \ \forall i, j;\ \omega \text{ free} \end{array} \tag{2.3}$$

where $\nu = (\nu_1, \nu_2, \ldots, \nu_m)^{\mathrm{T}}$ and $\mu = (\mu_1, \mu_2, \ldots, \mu_s)^{\mathrm{T}}$ are the vectors of input and output weights, or *multipliers*, and ω may be interpreted as a measure of scale efficiency along *one* facet of the frontier (see Banker and Thrall, 1992). The primal and dual programs (2.2) and (2.3) are commonly referred to as the *envelopment* and the *multiplier* problems, respectively.

The technology represented by (2.1) can be restricted to constant returns to scale by removing the requirement that the sum of the λs be equal to one. This leads to programs similar to (2.2), with the last constraint removed, and (2.3) without the variable ω. The resulting DEA model is usually referred to as the CCR model (Charnes et al., 1978). Other models and extensions can be found elsewhere (*c.f.* Ali and Seiford (1993), Charnes et al., 1995).

The BCC model constructs a piecewise linear envelopment surface (in $\Re_+^{m+s}$) from portions of the supporting hyperplanes that form the facets of the convex hull defined by the DMUs. The optimal multipliers in (2.3) are the coefficients of the linear equations that describe these facets; *i.e*. a solution (μ^*, ν^*) for a particular DMU yields *a* facet of the frontier that lies on the supporting hyperplane

$$\sum_{i=1}^{s} \mu_i^* y_i - \sum_{j=1}^{m} \nu_j^* x_j + \omega = 0 \tag{2.4}$$

However, the optimal multipliers at a given point (*e.g*. an extreme efficient point) may not be unique and several supporting hyperplanes may intersect at this point (see Figure 1 for a simple 2-D example). Thus, we cannot use the multipliers directly, without further consideration, as input and output “shadow” prices or, in the case of ω, as an indicator of returns to scale.

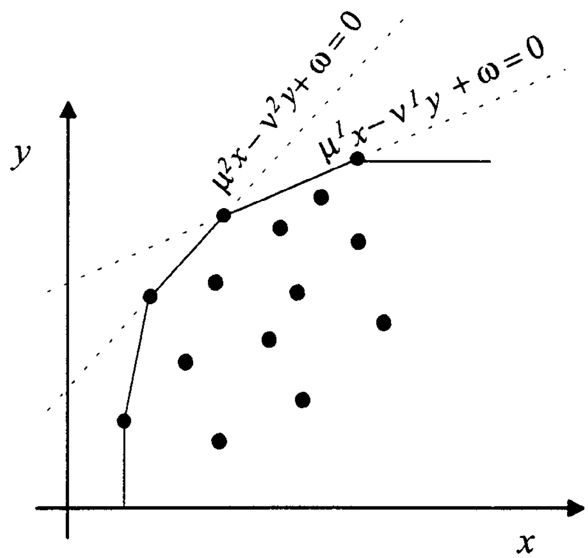

Figure 1. 2-D envelopment surface.

3. Marginal Rates & Partial Derivatives on the Frontier

Consider a general process, where a vector of outputs $\mathbf{y} = (y_1, y_2, \ldots, y_s) \in \Re_+^s$ is produced by a vector of inputs $\mathbf{x} = (x_1, x_2, \ldots, x_m) \in \Re_+^m$. To simplify the developments, let us write the inputs and outputs as a single vector of *throughputs* $\mathbf{z} \equiv (\mathbf{y}, -\mathbf{x})^{\mathrm{T}}$. We assume that the production frontier, given by the graph $\{\mathbf{z} : F(\mathbf{z}) = 0\}$, is continuous, piecewise differentiable, concave, and monotonic (there is free disposability of inputs and outputs). Note that frontiers constructed with DEA satisfy in general these conditions. The more familiar case in microeconomics textbooks of a single input y, with a production function $y = f(\mathbf{x})$, is simply a particular case with $F = y - f(\mathbf{x})$.

We refer to the partial derivative

$$\mathrm{MR}_{ij}(\mathbf{z}_o) \equiv \left.\frac{\partial z_i}{\partial z_j}\right|_{\mathbf{z}_o} = -\left.\frac{\partial F/\partial z_j}{\partial F/\partial z_i}\right|_{\mathbf{z}_o}, i \neq j \tag{3.1}$$

as the *marginal rate* of throughput i to throughput j, at a point $\mathbf{z}_o = (\mathbf{y}_o, -\mathbf{x}_o)^{\mathrm{T}}$ on the frontier ($F(\mathbf{z}_o) = 0$). In simple terms, this gives the increase in throughput i that results when throughput j is increased by one unit, in order to maintain the levels of the rest constant. It is straightforward to see that $\mathrm{MR}_{ij} = \mathrm{MR}_{ji}^{-1}$ and that, by construction of the frontier, $\mathrm{MR}_{ij} \leq 0$ for all i, j. Also, it is important to note that, strictly speaking, the marginal rates are only defined when the function F has continuous first derivatives (*i.e.* when $F \in C^1$).

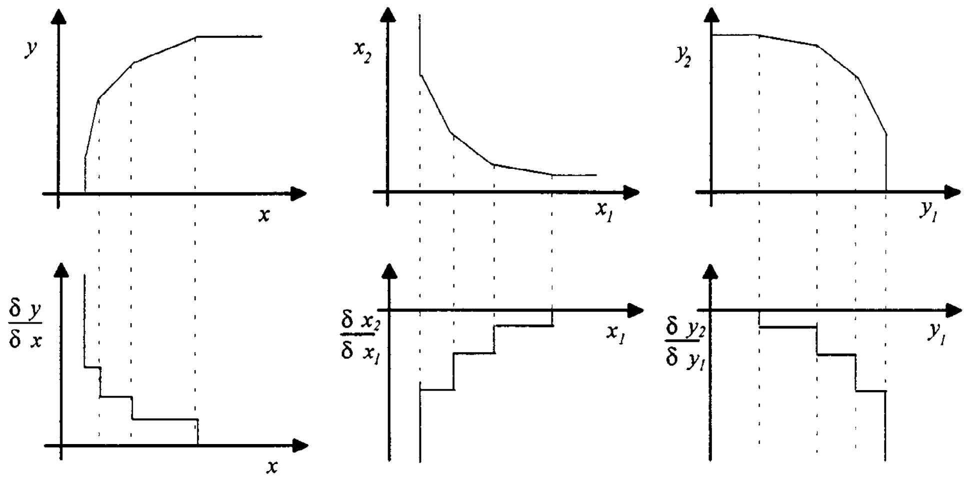

Figure 2. Behaviour of marginal rates in DEA.

When the throughputs correspond to two inputs, the rate (3.1) is usually referred to as the *marginal rate of technical substitution*, or simply the *technical rate of substitution*; when they correspond to two outputs, this partial derivative is called the *marginal rate of transformation*; and when z_i and z_j refer to an output and an input, respectively, it is often called the *marginal impact*, or *marginal productivity*; the inverse of the latter is the *marginal cost* (actually, because of the minus sign for the inputs in the definition of the throughput, this is the negative of the marginal productivity or cost).

Let us denote by $\chi^* = (\mu^*, \nu^*)$ the multipliers that describe a particular hyperplane that contains a facet of the DEA frontier. Then, (2.4) and (3.1) imply that the marginal rates at any interior point of this facet are simply given by the negative reciprocal of the ratio of two multipliers; that is,

$$\mathrm{MR}_{ij}(\mathbf{z}_o) \equiv \left.\frac{\partial z_i}{\partial z_j}\right|_{\mathbf{z}_o} = -\frac{\chi_j^*}{\chi_i^*} \tag{3.2}$$

Note that (3.2) is not strictly correct at the "edges" of the frontier since, as mentioned earlier, the multipliers are not unique and the partial derivative does not exist in a rigorous sense. This corresponds to the fact that the DEA envelopment surface is given by a piecewise linear relationship that is continuous, but only piecewise differentiable at these edges. Hence, as is depicted in Figure 2, the marginal rates are not uniquely defined there.

To deal with this problem, Charnes et al. (1985) suggested the use of the facet's average, or barycenter, to obtain quantitative estimates of the rates of change of outputs to inputs. Even then, the marginal rates are well defined only when the facet obtained from the solution of the linear program is full dimensional (*i.e.* it has dimension $m+s-1$). Quite frequently this is not the case and, therefore, the multipliers obtained may still not be unique. Moreover, this approach avoids the more interesting issue of characterizing the marginal rates *at* the extreme efficient points, which lie on the intersection of several facets. For this purpose, the method requires an algorithm that identifies *all* the facets that intersect at these points. Although such algorithms have recently become available (Fukuda, 1993), the number of facets, and the number of values for each trade-off, grows exponentially with the size of the problem. Moreover, such an approach does not directly discern whether movement in a particular direction along a given facet is feasible from a given point; *i.e*. whether a given rate on a facet is actually *achievable* from that particular point. Thus, a simpler and more productive approach may be to characterize the slopes at certain points neither as single nor multiple numbers but, more generally, in terms of their upper and lower bounds.

One formal solution to this problem is to define the marginal rates by the derivatives "to the right" and "to the left" at the particular point. At a point $\mathbf{z}_o \equiv (z_{1o}, \ldots, z_{(m+s)o})^{\mathrm{T}}$, these derivatives are given by:

$$\begin{aligned} \mathrm{MR}_{ij}^{+} &\equiv \left.\frac{\partial z_i}{\partial z_j}\right|_{\mathbf{z}_o^{+}} \\ &= \lim_{h\to 0^{+}} \frac{z_i(z_{1o}, \ldots, z_{jo}+h, \ldots, z_{(m+s)o}) - z_i(z_{1o}, \ldots, z_{jo}, \ldots, z_{(m+s)o})}{h} \end{aligned} \tag{3.3a}$$

$$\begin{aligned} \mathrm{MR}_{ij}^{-} &\equiv \left.\frac{\partial z_i}{\partial z_j}\right|_{\mathbf{z}_o^{-}} \\ &= \lim_{h\to 0^{-}} \frac{z_i(z_{1o}, \ldots, z_{jo}+h, \ldots, z_{(m+s)o}) - z_i(z_{1o}, \ldots, z_{jo}, \ldots, z_{(m+s)o})}{h} \end{aligned} \tag{3.3b}$$

where $z_i(z_{1o}, \ldots, z_{(i-1)o}, z_{(i+1)o}, \ldots, z_{(m+s)o})$ denotes the implicit function that gives the level of throughput i that places $\mathbf{z}_o$ on the frontier ($F(\mathbf{z}_o) = 0$), given the other throughputs. We refer to (3.3a) and (3.3b) as the *marginal rates to the right* and *marginal rates to the left*, respectively. Note that these rates are now well defined because of the piecewise construction of the DEA frontier.

We have used the vector of throughputs, $\mathbf{z}$, to simplify the notation and generalize the concept of the marginal rates to all combinations of inputs and outputs. A cautionary note is appropriate since these rates must be linked ultimately to those of the "real" inputs and outputs. This is briefly summarized in Table 1, where the equivalent rates for the original variables of the problem are given. For example, although the marginal rates (3.3) are always non-positive, the marginal productivities, when j corresponds to an input and i to an output, are always positive (*i.e*. $[\partial y_i/\partial x_{j-s}]^{\pm} \geq 0$, $1 \leq i \leq s$, $s+1 \leq j \leq m+s$). Furthermore, in this case, a derivative on the right ($+$), corresponds to one on the left ($-$) in the original variables (*i.e*. $(\partial z_i/\partial z_j)^{+} = (\partial y_i/\partial x_{j-s})^{-}$), and vice versa. In Table 1, $\mathrm{MRS}_{kl}^{\pm} = \left.\frac{\partial x_k}{\partial x_l}\right|_{x_l^{\pm}}$ denotes the marginal rates of technical substitution; $\mathrm{MRT}_{kl}^{\pm} = \left.\frac{\partial y_k}{\partial y_l}\right|_{y_l^{\pm}}$ the marginal rates of transformation; $\mathrm{MP}_{kl}^{\pm} = \left.\frac{\partial y_k}{\partial x_l}\right|_{x_l^{\pm}}$ the marginal productivities; and $\mathrm{MC}_{kl}^{\pm} = \left.\frac{\partial x_k}{\partial y_l}\right|_{y_l^{\pm}}$ the marginal costs.

Table 1. Marginal rates with inputs and outputs.

Throughputs	i input j input	i output j output	i output j input	i input j output
$\mathrm{MR}^+_{ij} =$	$\mathrm{MRS}^-_{(i-s)(j-s)}$	MRT^+_{ij}	$-\mathrm{MP}^-_{i(j-s)}$	$-\mathrm{MC}^+_{(i-s)j}$
$\mathrm{MR}^-_{ij} =$	$\mathrm{MRS}^+_{(i-s)(j-s)}$	MRT^-_{ij}	$-\mathrm{MP}^+_{i(j-s)}$	$-\mathrm{MC}^-_{(i-s)j}$

4. Computation of Marginal Rates

We introduce two computationally attractive approaches for obtaining the marginal rates outlined above: *(i)* computation of *minimum and maximum ratios* of multipliers (and its dual equivalent *finite differences*), and *(ii) modified simplex tableau* method. These approaches can be seen mostly as more general applications of similar developments in the DEA literature to handle returns to scale (RTS). In particular, the computation of "maximum and minimum ratios" is similar to the method suggested by Banker and Thrall (1992) to compute RTS intervals and, thus, overcome the problems in the original BCC paper (Banker et al., 1984). The finite differences method presents similar ideas to the approach for representing RTS in a recent paper by Golany and Yu (1994). These methods require the solution of two linear programs for the marginal rates at a given point on the frontier. The simplex tableau method is a more general application of the ingenious algorithm recently presented by Hackman et al. (1994) to characterize RTS and to represent a *(specific)* two-dimensional section of the production possibility set.

To simplify the presentation, we explicitly focus on a BCC frontier (*i.e*. variable returns to scale). However, the main concepts are valid for all DEA models, and the results can be readily generalized for each particular case.

4.1. Max-min Ratios and Finite Differences

Based on (3.2) and Figure 2, it seems that a reasonable way to obtain the marginal rates (3.3), at a given point on the frontier, is to compute the minimum and maximum ratios of multipliers at that point. Given the construction of the frontier and the definition of the throughput vector, the (negative of the) minimum and maximum multiplier ratios will correspond, respectively, to the marginal rate to the left and the marginal rate to the right. For a particular point $\mathbf{z}_0$ on the frontier, these ratios can be computed with the following two fractional programs:

$$
\begin{aligned}
\mathrm{MR}^+_{ij} = -\max_{\chi} \frac{\chi_j}{\chi_i} \qquad & \left(\mathrm{MR}^-_{ij} = -\min_{\chi} \frac{\chi_j}{\chi_i}\right) \\
\text{s.t. } \mathbf{z}_0^{\mathrm{T}}\chi + \omega &= 0 \\
\mathbf{z}_k^{\mathrm{T}}\chi + \omega &\le 0 \quad \forall k \in E \\
\chi &\ge 0
\end{aligned}
\tag{4.1}
$$

where $E = \{k : \text{DMU}_k \text{ is efficient}\}$, the set of indexes that correspond to DMUs on the efficient frontier (excluding the point under observation). It is only necessary to include the efficient DMUs, since the constraints that come from inefficient DMUs will not be binding.[1] The objective function in (4.1) minimizes (or maximizes) the possible marginal rates. The first constraint ensures that the point $\mathbf{z}_0$ remains on the frontier, while the second set of constraints forbids a choice of weights that would make any other DMU lie beyond the frontier (*i.e*. have an efficiency greater than unity). By introducing the change of variables $\rho_l = \chi_l/\chi_i$ $(l = 1, \ldots, m+s)$, $\omega^i = \omega/\chi_i$ (hence $\rho_i = 1$) , programs (4.1) are transformed into the corresponding linear programs:

$$
\begin{aligned}
\text{MR}_{ij}^{+} = -\max_{\rho,\omega^i} \rho_j \qquad & \left(\text{MR}_{ij}^{-} = -\min_{\rho,\omega^i} \rho_j\right) \\
& \rho_i = 1 \\
\text{s.t. } \mathbf{z}_0^{\mathrm{T}}\rho \; + \; & \omega^i = 0 \\
\mathbf{z}_k^{\mathrm{T}}\rho \; + \; & \omega^i \le 0 \quad \forall k \in E \\
& \rho \ge 0
\end{aligned}
\tag{4.2}
$$

As can be seen in Figure 2, the minimum value for a marginal rate may be unbounded. This occurs when the multiplier in the denominator vanishes, and should not cause any problems since any linear programming package will detect it. Accordingly, when the maximum ratio is zero, the multiplier in the numerator vanishes. Both of these cases correspond to points on the "endings" of the frontier, where the isoquant becomes parallel to one of the axes.

The dual formulation of (4.2) has a very interesting interpretation that relates to the finite difference approximation of the rates. Finite difference approximations of partial derivatives have been widely applied in numerical methods for solving partial differential equations (*c.f.* Ames, 1992), nonlinear equations and optimization (*c.f.* Dennis and Schnabel, 1983). In particular, first order approximations of a partial derivative of a (sufficiently continuous) function $f\colon \Re^m \to \Re$ are given by

$$
\frac{\partial f}{\partial x_j}(\mathbf{x}_0) \approx \frac{f(\mathbf{x}_0 + h\mathbf{e}_j) - f(\mathbf{x}_0)}{h}
\tag{4.3a}
$$

$$
\frac{\partial f}{\partial x_j}(\mathbf{x}_0) \approx \frac{f(\mathbf{x}_0 - h\mathbf{e}_j) - f(\mathbf{x}_0)}{-h}
\tag{4.3b}
$$

where $\mathbf{e}_j$ denotes the j-th unit vector (*i.e*. a vector of zeros with a one in the j-th position) and h is a (small) finite positive number.[2] We call (4.3a) and (4.3b) the *forward difference* and *backward difference*, respectively. By noticing the similarity between (3.3) and (4.3), it becomes apparent that the forward and backward differences can be used to obtain the marginal rates on the right and on the left for our particular case where the frontier has only piecewise continuous derivatives. And, since the surface is piecewise linear, the first order approximations (4.3) give the *exact* slopes, provided the step h is small enough to remain on an adjacent facet.

Using these ideas, an intuitive procedure to compute the marginal rates at a point $\mathbf{z}_0$ on the frontier can be formulated in three steps: *(i)* define a *small* step h; *(ii)* obtain the implicit function $z_{io}^{\pm}(z_{1o}, \ldots, z_{jo} \pm h, \ldots, z_{(m+s)o})$ that results from either increasing or decreasing the j-th throughput by h; and *(iii)* compute the finite differences. Step *(ii)* requires the solution of the following linear programs, to obtain $z_{io}^{\pm}$:

$$
\begin{aligned}
\max_{\lambda_o,\lambda,z^{\pm}} \ & z_{io}^{\pm} \\
\text{s.t. } \mathbf{z}_o\lambda_o + \mathbf{Z}\lambda \ &\geq \ \mathbf{z}_o^{\pm} \\
\lambda_o + \mathbf{e}^{\mathrm{T}}\lambda \ &= \ 1 \\
z_{lo}^{\pm} \ &= \ z_{lo} \qquad l \neq i, j \\
z_{jo}^{\pm} \ &= \ z_{jo} \pm h \\
\lambda_o, \lambda_k \ &\geq \ 0, \quad k = 1, \ldots, n_E \\
z_{lo}^{\pm} \geq 0, l = 1, \ldots s;\ z_{lo}^{\pm} \ &\leq \ 0, \quad l = s+1, \ldots, m+s
\end{aligned}
\tag{4.4}
$$

where $\mathbf{Z}$ is the matrix of throughputs with columns $\mathbf{z}_k$, $k \in E$, and n_E is the number of efficient DMUs (or indices in E). Program (4.4) solves for the i-th component in the *new* throughput vector $\mathbf{z}_0^{\pm}$, such that it remains on the frontier when the j-th component is increased, or decreased, by a small quantity h, and all the other components are maintained constant. The marginal rates can then be calculated as

$$
\mathrm{MR}_{ij}^{+}(\mathbf{z}_0) = \frac{z_{i_0}^{+} - z_{i_0}}{h} \quad \text{and} \quad \mathrm{MR}_{ij}^{-}(\mathbf{z}_0) = \frac{z_{i_0}^{-} - z_{i_0}}{-h}
\tag{4.5}
$$

where, abusing the notation, $z_{io}^{\pm}$ is used to designate the optimum in programs (4.4).

Although the formulation (4.4)–(4.5) using finite differences is intuitively appealing, it requires the explicit choice of the step h and may not be the best alternative for practical computations. Here, we show how this program can be transformed to one that is more computationally attractive, and how it relates to the max-min ratio LPs (4.2). Since the max-min programs (4.2) are on the "multiplier side" of a DEA formulation, and the linear programs we must solve for the finite differences are on the "envelopment side", it seems reasonable to seek the relationship between the two through their corresponding dual programs.

Let us start by simplifying the program (4.4) and incorporating the expressions for the marginal rates (4.5). For this purpose, we write the new vector $\mathbf{z}_o^{\pm}$ as $\mathbf{z}_o^{\pm} = \mathbf{z}_o + h\mathbf{e}_j + \Delta^{\pm}\mathbf{e}_i$ where $\Delta^{\pm}$ is the resulting increment in the i-th component of $\mathbf{z}_o$ from an increase of $\pm h$ in its j-th component; note that $\Delta^{\pm}$ can be positive or negative. Then, $\mathrm{MR}_{ij}^{\pm} = \frac{\Delta^{\pm}}{\pm h}$ and, substituting in (4.4), we obtain

$$
\begin{aligned}
\max_{\Delta^{\pm},\lambda_o,\lambda} \ & \Delta^{\pm} \\
\text{s.t. } -\mathbf{e}_i\Delta^{\pm} \ + \ & \mathbf{z}_o(\lambda_o - 1) + \mathbf{Z}\lambda \geq \pm h\mathbf{e}_j \\
& (\lambda_o - 1) + \mathbf{e}^{T}\lambda = 0 \\
\lambda_o, \lambda_k \geq 0, \quad k \ = \ & 1, \ldots, n_E;\ \ \Delta^{\pm} \text{ free}
\end{aligned}
\tag{4.6}
$$

The optimal value that corresponds to the objective function in (4.4) is now $z_{io}^{\pm} = z_{io} \pm \Delta^{\pm}$. If we now divide every constraint by h, and make the following substitutions

$$\varphi^{\pm} = -\frac{\Delta^{\pm}}{h}, \quad \lambda_o' = \frac{\lambda_o - 1}{h}, \quad \lambda' = \frac{1}{h}\lambda \tag{4.7}$$

we obtain the following linear programs:

$$\begin{aligned} \mathrm{MR}_{ij}^{\pm} = \pm \max_{\varphi^{\pm},\lambda_0',\lambda'} & -\varphi^{\pm} \\ \text{s.t. } \mathbf{e}_i\varphi^{\pm} \ + \ & \mathbf{z}_0\,\lambda_0' + \mathbf{Z}\,\lambda' \geq \pm\mathbf{e}_j \\ & \lambda_0' + \mathbf{e}^{\mathrm{T}}\lambda' = 0 \\ \lambda_k' \geq 0, \quad k \ + \ 1, \ldots, n_E; \ & \varphi^{\pm}, \lambda_o' \text{ free} \end{aligned} \tag{4.8}$$

This program is simpler than (4.4), it computes the rates directly and, most of all, it does not require the explicit definition of the step h. Furthermore, it can be readily shown that program (4.8) is the dual program of the max-min ratio program (4.2).

4.2. Modified Simplex Tableau

Hackman et al. (1994) introduced an ingenious algorithm to characterize the two-dimensional section of the frontier representing the amounts by which the inputs and outputs of a DMU could be *scaled* and still lie in the production possibility set. Hence, this algorithm, which closely resembles the simplex method, also yields technical and scale efficiencies in DEA. In this section, we describe a more general application of this algorithm to characterize *any* two-dimensional section of a multi-dimensional DEA production frontier. Accordingly, the algorithm readily gives the marginal rates at certain *levels* of the other variables. The method can be used also to obtain plots of level curves. This results in a better geometrical picture of the production possibilities and a deeper understanding of the implicit trade-offs.

Consider the general case where we want to generate the two-dimensional section, z_i *vs.* z_j, of the envelopment surface at the level of a point $\mathbf{z}_o$ (an efficient or inefficient DMU); *i.e.* we want to obtain the graph $\{(z_j, z_i) : F(z_{1o}, \ldots, z_j, \ldots, z_i, \ldots, z_{(m+s)o}) = 0\}$. Clearly, the trade-offs between the two variables at this level are given by the slopes of the section. Let us focus on the envelopment side of the DEA formulation and generate a simplex tableau that now also includes z_j and z_i as columns. Then, we consider the set of constraints now given by

$$\begin{aligned} -\mathbf{e}_i z_i - \mathbf{e}_j z_j + \mathbf{z}_o\lambda_o + \mathbf{Z}\lambda - \mathbf{I}\,\mathbf{s} \ &= \ \mathbf{z}_o - \mathbf{e}_i z_{io} - \mathbf{e}_j z_{jo} \\ \lambda_o + \mathbf{e}^{\mathrm{T}}\lambda \qquad &= \ 1 \\ \lambda_o, \lambda_k \ &\geq \ 0, k = 1, \ldots, n_E \\ z_l \ &\geq \ 0 \text{ if } l \leq s, z_l \leq 0 \text{ if } l > s, l = i, j \end{aligned} \tag{4.9}$$

where $\mathbf{s}$ is the vector of slacks, and $\mathbf{I}$ is the identity matrix of dimension $(m + s)$; note that we have explicitly included the observation $\mathbf{z}_o$ in the reference set. A feasible solution to (4.9) is given by $\{z_i = z_{io}, z_j = z_{jo}, \lambda_o = 1, \lambda = \mathbf{0}, s_l = 0\,\forall\, l\}$. Then, we can construct

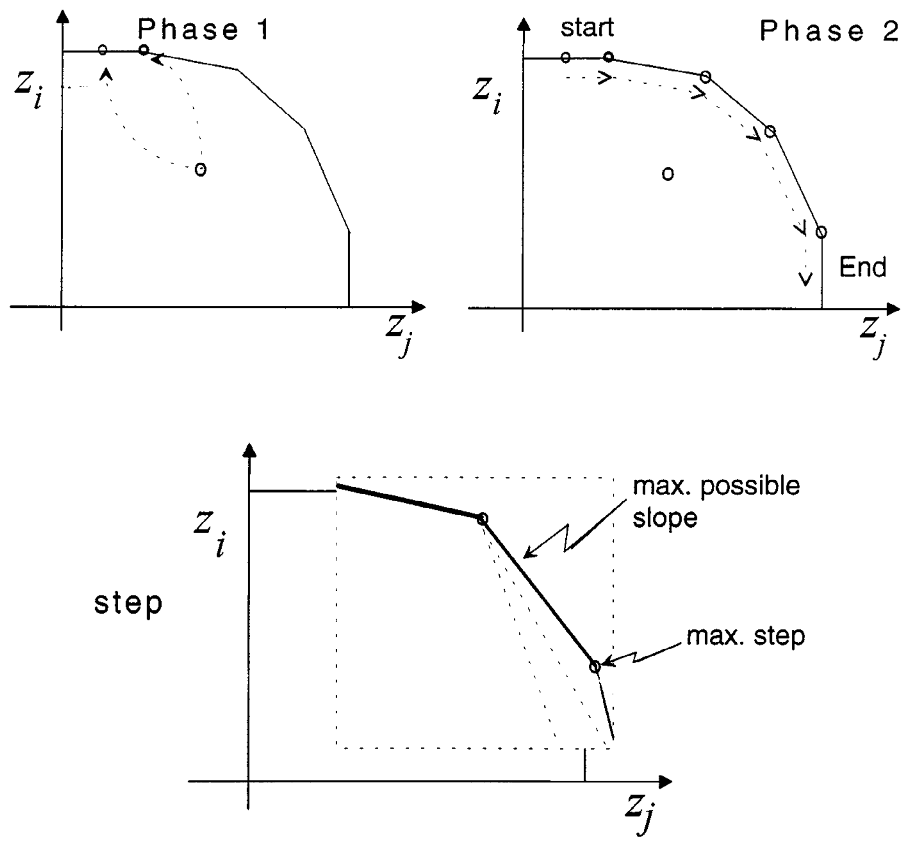

Figure 3. Modified simplex tableau algorithm.

a simple initial feasible basis $\mathbf{B}_o$ with the variables $(z_i, z_j, \lambda_o, s_l\ \forall l \neq i, j)$ to generate the initial simplex tableau.

The algorithm consists of two phases (see Figure 3): in phase 1, we find the uppermost point in the section by maximizing z_i. This is done using standard pivoting. As a result we obtain the "first" vertex of the section, $\xi^o \equiv (z_j^o, z_i^o) = (z_j, z_i)\,|_{\text{end of Phase 1}}$. As is shown in Figure 3, if inefficient points are included, ξ^o may not be a "true" vertex, but a point in a segment of the isoquant parallel to the z_j-th axis. This, however, does not cause any problems (see Section 5.1).

Similar to the simplex method, phase 2 now consists of finding a sequence of pivots that correspond to all the adjacent vertices. Thus, by consecutively pivoting to the right (by increasing z_j) we obtain all the vertices $(\xi_1, \xi_2, \ldots, \xi_p)$ and, at each of these, the marginal rates to the right and to the left. Let us describe phase 2 in more detail. For this purpose, we denote by B the set of basic variables, NB the set of non-basic variables, A_{rq} the entry in the r-th row and q-th column of the simplex tableau, $\mathbf{u} \equiv (\lambda_o, \lambda, \mathbf{s})$, and (z_i^k, z_j^k) the values of (z_i, z_j) *after* the k-th step of the tableau. If, for simplicity, we further use the subindices i and j to denote the rows corresponding to the variables z_i and z_j, respectively, the constraints associated with these rows

$$z_i + \sum_{l \in NB} A_{il} u_l = z_i^k$$

after the k-th step are then given by

$$z_j + \sum_{l \in NB} A_{jl} u_l = z_j^k \tag{4.10}$$

It is straightforward to see that, if the q-th variable enters the basis in the next step, the rate of change of z_i with respect to z_j at this step, MR_{ij}^{k+1}, is given by

$$\mathrm{MR}_{ij}^{k+1} \equiv \left.\frac{\partial z_i}{\partial z_j}\right|^{k+1} \equiv \frac{z_i^{k+1} - z_i^k}{z_j^{k+1} - z_j^k} = \frac{A_{iq}}{A_{jq}} \tag{4.11}$$

When pivoting to the right, we want to increase z_j; hence, we can see from (4.10) that column candidates to enter the basis should have a negative entry in the j-th row. Since an increase in z_j will result in a decrease in z_i, the entry in the i-th row will be positive, unless the original point lies at the ends of the isoquant (where it may be zero). Then, in order to move along a line which "envelops" (or, more technically, which supports) the data, we choose the column that gives the highest (*i.e*. the least negative) ratio (4.11). If we were to move along a steeper slope, we would remain within the feasible region, but not on the frontier (see Figure 3). Similarly, pivoting to the left would entail the choice of the minimum ratio (4.11) with a positive entry for the j-th column. More formally, this leads to the following rules for choosing the pivot column, pc:

$$\begin{aligned} pc = q : \frac{A_{iq}}{A_{jq}} &= \max_{l \in NB} \left\{ \frac{A_{il}}{A_{jl}}, A_{jl} < 0 \ (A_{il} \geq 0) \right\} \text{ pivot right} \\ pc = q : \frac{A_{iq}}{A_{jq}} &= \min_{l \in NB} \left\{ \frac{A_{il}}{A_{jl}}, A_{jl} > 0 \ (A_{il} \leq 0) \right\} \text{ pivot left} \end{aligned} \tag{4.12}$$

In the absence of degeneracy, we expect z_j to increase at each iteration when we pivot to the right. Furthermore, at each step, the ratio A_{iq}/A_{jq} in (4.12) gives the corresponding marginal rate. When we reach the last vertex, the isoquant becomes parallel to the z_i-th axis and the rate of change goes to minus infinity.[3] However, in the presence of degeneracy, this ratio cannot be directly interpreted as an *actual, realizable*, rate of change at this level. Actual marginal rates are only obtained once we "physically" move from the point (z_j^k, z_i^k); *i.e*. when $(z_j^{k+1}, z_i^{k+1}) \neq (z_j^k, z_i^k)$. However, because of the construction of the algorithm, the ratio A_{iq}/A_{jq} *always* correspond to the rate on *one* facet of the frontier at the vertex, even though movement on that particular facet in the direction of increasing z_j, from the current vertex, may not be feasible. This will be shown more clearly with the example in

Section 5.3. The pivot row, pr, can then be chosen in the usual way:

$$pr = r : \frac{b_r}{A_{r\ pc}} = \min_l \left\{ \frac{b_l}{A_{l\ pc}}, A_{l\ pc} > 0 \right\} \tag{4.13}$$

where b_l denotes the l-th entry on the right hand side of the simplex tableau. This way we restrict how far we can move without violating any of the constraints.

By following these pivots, we proceed to move forward on the frontier from vertex to vertex and, thus, effectively generate the whole 2-D section of the isoquant at that level. Note that at each step, the basic solution corresponds to a vertex of the 2-D section of the isoquant. Accordingly, at each (non-degenerate) step we also obtain the corresponding marginal rate from moving in that direction. When several columns have the same ratio, the resulting vertices are co-linear and we may or may not choose to generate all these vertices.

A mathematically formal validation of the modified simplex tableau algorithm is given by Hackman et al. (1994). Instead, in this section, we have attempted to explain the rationale of the algorithm using quite simple and intuitive arguments. Section 6 presents a general characterization of derivatives on the frontier where the marginal rates and returns to scale are special cases. This results also in the generation of more general 2-D sections.

One can argue that the piecewise linear DEA frontier is simply an approximation of a "real" smooth, multidimensional, concave, surface. Then, the marginal rates to the right and to the left give bounds for a "true", unique, marginal rate. A simple estimate of this single number is given by the average of the two rates. Perhaps, a more appropriate estimate is a weighted average, where the weights reflect the lengths of the linear segments of the section that intersect at this point. Thus, if we denote by L^+ and L^- the lengths of the segments to the right and left of the point respectively, an estimate of a marginal rate can be given by $\mathrm{MR}_{ij} \approx \frac{1}{L^+ + L^-}(L^+\mathrm{MR}_{ij}^- + L^-\mathrm{MR}_{ij}^+)$. Note that the weight accorded to each rate corresponds to the length of the opposite segment; this way, we give a higher weight to the rate that corresponds to the shortest segment. Note that this estimate, however, is very sensitive to the measurement units and to the type of norm used. An alternative to this approach is to fit a smooth function (for example using C^1 splines) through the vertices of the 2-D section obtained with the modified simplex algorithm. The rate at any point can then be computed from the derivative of this function.

5. Examples

To understand the previous algorithm better, we first examine a simple example with two inputs and one (constant) output in order to show how the modified simplex algorithm replicates the plot of a conventional two-input isoquant (for a certain level of output). Then, we employ one example from the literature to show how one can apply the concepts and techniques introduced in the previous sections to give a more rigorous, and meaningful, treatment of marginal rates. A third, more complex, example is used to show the wider applicability of the methodology and the usefulness of the visualization capabilities that result from it.

5.1. 2-Input/1-Output Example

Consider the data in Figure 4 for eight DMUs that produce a single output using two inputs. We proceed to construct the section of the isoquant defined by the two inputs, at a constant level of $y = 1$. Let us start with the initial point $\mathbf{z}_o = (-3, -2, 1)$. From the figure, we can see that the range of marginal rates of technical substitution of input 1 to input 2 at this point is given by $\text{MRS}_{12}^{-} = -2$ and $\text{MRTS}_{12}^{+} = -0.5$. For simplicity, using a constant returns to scale model, the initial matrix of constraints and right hand side (4.9) for this particular example is given by

Matrix of contraints & RHS

x1	x2	λ0	λ1	λ2	λ3	λ4	λ5	λ6	λ7	λ8	s1	s2	sy	RHS
−1	0	−3	−1	−1	−2	−3	−5	−7	−4	−3	−1	0	0	0
0	−1	−2	−8	−6	−4	−2	−1	−1	−4	−7	0	−1	0	0
0	0	1	1	1	1	1	1	1	1	1	0	0	−1	1

(5.1)

where (s1, s2, sy) denote the slacks in the two inputs and output. (We have retained the column for DMU 4, although it is not necessary since it is the same as that for DMU_O. The inefficient DMUs are also retained in the table to show that they never enter the basis in phase 2.) Using the first three variables as a basis, the initial simplex tableau is

Initial Tableau

Basis	x1	x2	λ0	λ1	λ2	λ3	λ4	λ5	λ6	λ7	λ8	s1	s2	sy	RHS
x1	1	0	0	−2	−2	−1	0	2	4	1	0	1	0	3	−3
x2	0	1	0	6	4	2	0	−1	−1	2	5	0	1	2	−2
λ0	0	0	1	1	1	1	1	1	1	1	1	0	0	−1	1

(5.2)

As described earlier, in step 1 we attempt to maximize x_1; hence, we proceed to look for the column with the most negative entry in the first row, and choose the pivot row in the regular way. The resulting pivot is column 4 and row 3 (λ_1 enters the basis while λ_o leaves). Note that column 5 might have been chosen as well. Then, the pivoting step results in the following table:

Phase 1 (step 1)

Basis	x1	x2	λ0	λ1	λ2	λ3	λ4	λ5	λ6	λ7	λ8	s1	s2	sy	RHS
x1	1	0	2	0	0	1	2	4	6	3	2	1	0	1	−1
x2	0	1	−6	0	−2	−4	−6	−7	−7	−4	−1	0	1	8	−8
λ1	0	0	1	1	1	1	1	1	1	1	1	0	0	−1	1

(5.3)

Since all the entries in the first row are non-negative, the first phase is complete and the maximum x_1 has been achieved.

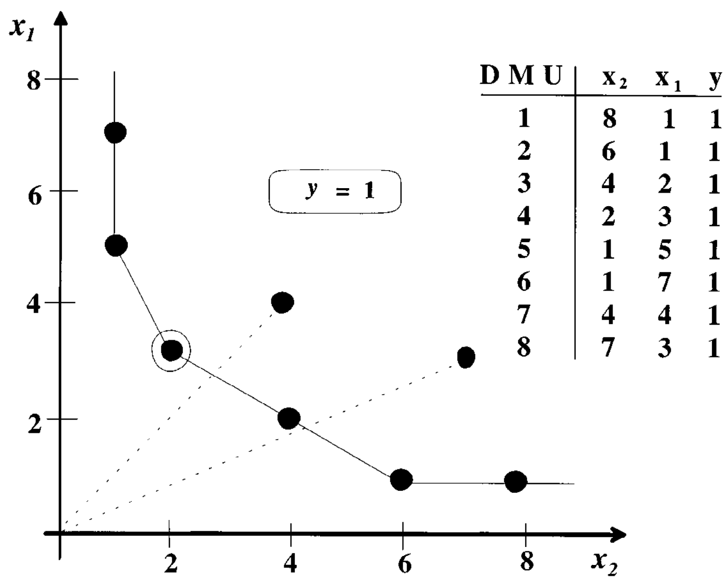

Figure 4. 3-D example.

We proceed to generate the whole section by pivoting to the right (*i.e.* seeking to maximize x_2 at each step). Candidates to enter the basis are all the columns with a negative entry in the second row (λ_o, λ_2 to λ_8). Then, following (4.12), λ_2 enters since it has the highest ratio of the entry in the first row to that of the second one; the marginal rate (to the right) at this point is equal to zero; and, following (4.13), λ_1 leaves the basis. The resulting tableau, as well as those for all consecutive steps are presented below. At each step, we also give the marginal rate (to the right) that results from pivoting in that tableau.

Phase 2

Step 1 ($\mathbf{MR^0 = 0}$)

Basis	x1	x2	λ0	λ1	λ2	λ3	λ4	λ5	λ6	λ7	λ8	s1	s2	sy	RHS
x1	1	0	2	0	0	1	2	4	6	3	2	1	0	1	−1
x2	0	1	−4	2	0	−2	−4	−5	−5	−2	1	0	1	6	−6
λ2	0	0	1	1	1	1	1	1	1	1	1	0	0	−1	1

Step 2 $(\mathbf{MR^1 = 0.5})$

Basis	x1	x2	λ0	λ1	λ2	λ3	λ4	λ5	λ6	λ7	λ8	s1	s2	sy	RHS
x1	1	0	0	−2	−2	−1	0	2	4	1	0	1	0	3	−3
x2	0	1	0	6	4	2	0	−1	−1	2	5	0	1	2	−2
λ4	0	0	1	1	1	1	1	1	1	1	1	0	0	−1	1

Step 2 $(\mathbf{MR^2 = 2.0})$

Basis	x1	x2	λ0	λ1	λ2	λ3	λ4	λ5	λ6	λ7	λ8	s1	s2	sy	RHS
x1	1	0	−2	−4	−4	−3	−2	0	2	−1	−2	1	0	5	−5
x2	0	1	1	7	5	3	1	0	0	3	6	0	1	1	−1
λ5	0	0	1	1	1	1	1	1	1	1	1	0	0	−1	1

End $(\mathbf{MR^3 = -inf})$

(5.4)

In the last table, it is not possible to have an increase in x_2 since all the entries in the second row are non-negative. At this point, the section becomes parallel to the x_1 axis and goes to minus infinity (thus, passing through DMU_6). Note also that, in step 2, λ_3 could have been chosen to enter the basis as well; this means that the corresponding vertices are co-linear. In this case, we have chosen to move to the vertex that is further away.

In this example, all the vertices are non degenerate and, hence, at each step we obtain the true marginal rates. Section 5.3 presents a more complex example when this is not the case.

Figure 5 further presents a level plot of the envelopment surface under constant returns to scale. This is simply the section generated with the repeated application of the algorithm for different levels of the output y.

5.2. Marginal Costs in Hospitals

Thanassoulis (1991) calculated DEA estimates of cost per unit output for 15 hypothetical hospitals, for a three output and one input production model. The single input was the total cost, while the outputs were the number of teaching units, regular patients, and severe patients. The data was originally generated by Sherman (1984) using a linear cost function and, for completeness, is reproduced in Table A.1 in the Appendix.

In Thanassoulis (1991), the marginal costs were basically estimated from *one* set of optimal weights of a standard DEA formulation. As is pointed out correctly in the paper, the multiplicity of the weights must be considered when analyzing the marginal costs, and not exclusively the set of marginal rates that arise from the first optimal solution. However, the study did not present a solution to this problem and only pointed to the inventive technique of constrained facet analysis (Bessent et al., 1988) to obtain positive weights for as many variables as possible. In Table 2, we give the marginal costs to the right and to the left (calculated with the techniques presented in Section 3) for each hospital; the value reported by Thanassoulis is given in parenthesis. For the inefficient DMUs (DMUs 8–15)

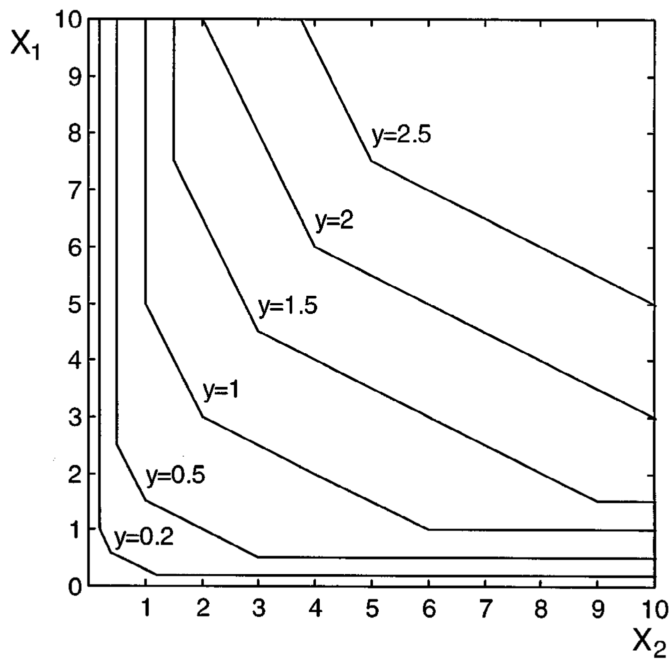

Figure 5. Input level curves.

the marginal costs are reported at their input radial projection on the frontier (as opposed to the efficient projection that also accounts for the possible output slacks). Hence, some rates are exactly zero. Had we used the efficient projection instead, at least one of the bounds would be non-zero for each rate and each DMU.

The benefits of the current formulation are evident even for this quite simple example. As is expected, an arbitrary solution of the DEA linear program with an extremal method, such as the simplex, yields only one of the bounds on the marginal rates (either to the right or to the left). By providing information on the full range of values that the rates can take at any particular point, the current methodology provides a more accurate representation of the costs. Moreover, when a rate goes to zero at one extreme of the isoquant, we are able to obtain the non-zero rate that corresponds to the efficient frontier, if it exists; *i.e*. the "last" slope of the efficient frontier before it ends (and the isoquant becomes parallel to one of the axis). The ranges of the marginal costs presented here give stronger and more rigorous support to the conclusions of the original paper by Thanassoulis: namely, that DEA can give better information on trade-offs than that obtained by the more widely used regression techniques.

Table 2. Marginal costs (cost per unit of output) for simulated hospitals.

Hospital	$/teaching unit	$/regular patient	$/severe patient
1	497–500 (496)	133.67–133.69 (133.60)	174.72–174.79 (177.71)
2	0–498 (0)	127.46–133.70 (127.94)	174.78–187.23 (187.92)
3	502–7183 (7161)	0–133.63 (0)	41.10–174.73 (41.04)
4	500–8005 (8000)	0–154.24 (0)	0–174.74 (0)
5	0–490 (0)	133.70–134.72 (133.99)	174.90–183.07 (183.46)
6	0–3174 (0)	0–133.68 (0)	174.73–238.21 (238)
7	0–3378 (0)	133.68–171.13 (172.41)	0–182.05 (0)
8	500–8005 (8000)	0–154.24 (0)	0–174.74 (0)
9	0–498 (0)	127.46–133.70 (127.94)	174.78–187.23 (187.92)
10	3378–3378 (3387)	154.24–154.24 (154.6)	0–0 (0)
11	0–0 (0)	134.72–134.72 (135)	182.05–182.05 (183)
12	500–500 (500)	133.67–133.67 (134.33)	174.74–174.74 (175.6)
13	0–0 (0)	133.70–133.70 (133.83)	183.07–183.07 (183.23)
14	497–500 (495)	133.67–133.69 (134)	174.72–174.79 (175.21)
15	500–500 (500)	133.67–133.67 (134.33)	174.74–174.74 (175.6)

5.3. 2-D Level Plots in Software Production

Example 5.1 involved a fictitious, ideal, 3-D production possibility set that showed no degeneracy. In this section, we show more realistic 2-D level plots from a real data set of 15 software projects in a large Canadian bank (Paradi et al., 1995). The data for these projects and their input efficiency scores are given in Table A.2 in the Appendix. The production model consists of one input, project cost, and three outputs: size (functionality) measured in function points; duration, measured in days; and quality, measured in rework-hours. For the purpose of the analysis, the last two outputs have been transformed by subtracting the original value (days or rework-hours) from the maximum of each within the sample, so that the new outputs are to be maximized (see Paradi et al. (1995) for a detailed discussion of software production and the models). Commonly, software management has had a special interest in the cost/functionality, quality/functionality and cost/time trade-offs. However, it can be quite misleading to analyze the trade-offs in isolation, without considering the other variables simultaneously, as is commonly done with performance ratios. This problem can be overcome quite effectively with the methods presented in section 4.

We start by investigating the quality/functionality trade-offs at the level of each efficient DMU (DMUs 5, 6, 8, 11, 15) under a variable returns to scale (VRS) technology. We use the modified simplex algorithm of section 4.4 to generate the 2-D sections that correspond to each of these DMUs. Table 3 shows, for each DMU, the sequence of vertices generated by the algorithm and the marginal rates of transformation at each vertex. Note that degeneracy appeared in all but one case (that of DMU 6). Thus, for example, when solving for DMU 8, we find first the uppermost vertex $\text{v1} = (64.74, 1657.8)$. At this vertex, the marginal rate to the left is, obviously, zero; this can also be checked by pivoting to the left at this point. We then proceed with the second phase and pivot to the right. This phase consisted of four pivoting steps. In the first step, the marginal rate to the right at v1 is found to be -0.294; this means that, at this point, an increase of one function point (FP) shall result in a decrease in quality of 0.294, for the same cost and duration (*i.e.* in an increase of 0.294 rework hours, because of the data transformation). This is an achievable rate since the next vertex is different $\text{v2} = (509, 1527)$. In turn, -0.294 is the marginal rate to the left of v2. The next pivoting step yields a rate of -0.337. However, this rate is not achievable since we have a degenerate solution at this point and there is no movement allowed in this direction, at this level (the new basic solution corresponds to the same vertex v2). This is also true in the third step where we get a non-achievable rate of -0.720. The last step shows no possibility of moving to the right (increasing the number of function points), and we obtain a marginal rate of minus infinity.

Note that, while the rates of -0.337 and -0.720 were not achievable at the level of DMU 8, they were achievable rates at the level of DMU 15. Indeed, as mentioned earlier, by construction of the algorithm, the slopes obtained at each pivoting step are always true rates of *one* facet where the vertex lies. However, they may not be achievable at the level of that particular vertex.

Figure 6a presents the quality/functionality data for the given projects. The lines define the sections of the isoquant at the level of each of the efficient DMUs under VRS. The slopes of the lines are the achievable trade-offs in Table 3. The plots of the 2-D sections provide a good, intuitive, graphical representation of the multi-dimensional frontier. Thus, for instance, although DMUs 5 and 6 seem to dominate all the others in these two variables, when the level of the other two factors are considered, we see the section against which each DMU must be compared. Similar plots can be drawn also at the levels of the inefficient DMUs.

Other interesting notions become evident from the figure. For example, the first vertex at the level of DMU 8 (64.74, 1657.8) corresponds to a convex combination of DMUs 5 and

Table 3. Vertices & marginal rates of transformation—quality/functionality—for efficient DMUs under VRS technology.

	Number of (true)	Vertices		Marginal Rates	
	Vertices	Function Points	Quality	to the left	to the right
DMU 5	1	169	1813	0	−0.368
		169	1813		−2.86
		169	1813		− inf
DMU 6	2	169	1813	0	−0.368
		609	1615	−0.368	− inf
DMU8	2	64.74	1657.8	0	−0.294
		509	1527	−0.294	−0.337
		509	1527		−0.720
		509	1527		− inf
DMU 11	1	334	1484	0	−0.337
		334	1484		−0.720
		334	1484		− inf
DMU15	3	32	1609	0	−0.294
		32	1609		−0.337
		362.93	1497.4	−0.337	−0.720
		367.88	1493.9	−0.720	− inf

15. Similarly, the second vertex at the level of DMU 15 (362.93, 1497.4) is a combination of DMUs 11 and 6, while the third one (367.88, 1493.9) is a combination of DMUs 11 and 8). This can also be obtained from the basic solutions at each step of the algorithm.

The results suggest that the observed (non-zero, finite) marginal rates of transformation between quality and functionality, at the efficient points, lie mostly in the range of (−0.337, −0.294) (only a small portion of the section at the level of DMU 15 showed a more negative rate of −0.720). Thus, for projects that lie within the range of these data, it is reasonable to expect an increase of about 0.3 rework hours per additional FP.

Figure 6b shows the quality/functionality level curves for the same DMUs under a constant returns to scale (CRS) technology. Only DMUs 5 and 11 are scale efficient. Also, we can see from the figure that DMUs 6 and 8 show slacks in quality while DMU 15 shows a slack in functionality. Note that the CRS production possibility set is much larger than the one under VRS at the level of DMUs 6 and 8.

The only non-zero marginal rate obtained from the CRS frontier is −16.22, which implies an increase in 16.22 rework hours/FP. The fact that this is a very unrealistic rate for this process at the level of these DMUs, supports the conclusion in Paradi et al. (1995) of the existence of variable returns to scale within the process.

As another example, Figure 7 gives the cost/functionality level plots for the efficient DMUs under VRS. In this case, we are analyzing the marginal costs of functionality and, hence, the frontier bounds the data from the south-east. The observed finite, non-zero, trade-offs are in the range of 606 $/FP to 6790 $/FP (the solution for DMUs 11 and 15 also

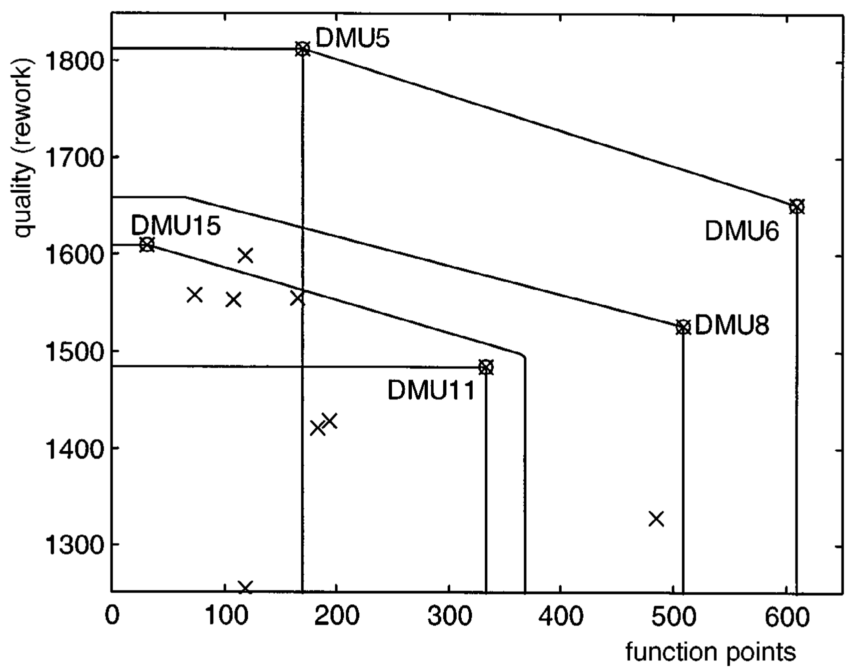

Figure 6a. Quality/functionality trade-offs − VRS.

gave a non-achievable rate of 15500 $/FP). More detailed information on the marginal rates is given Figure 8, where the cost per FP is plotted against the number of FPs at the level of each project.

6. Generalizations

The treatment of marginal rates introduced in Section 4 can be seen as a particular case of a more general formulation via directional derivatives on the efficient frontier. The standard definition of the directional derivative of a function $f(\mathbf{x})$ in the direction of a unit vector $\mathbf{v}$ is given by

$$\frac{\partial f}{\partial \mathbf{v}} = \frac{d}{dt} f(\mathbf{x} + t\mathbf{v})\bigg|_{t=0} = \lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x})}{t} \tag{6.1}$$

As done in Section 2, consider now the general production frontier $\{\mathbf{z} : F(\mathbf{z}) = 0\}$. Then, for a pair of vectors $\mathbf{u}$ and $\mathbf{v}$, we are interested in assessing the change of the throughputs in the direction of $\mathbf{v}$ that occurs with a small change of the throughput vector in the direction of $\mathbf{u}$, while remaining on the frontier. Then, if we denote by α the change in the $\mathbf{v}$-direction that results from moving β units in the $\mathbf{u}$-direction, with $F(\mathbf{z}+\alpha\mathbf{v}+\beta\mathbf{u}) = 0$, we can define

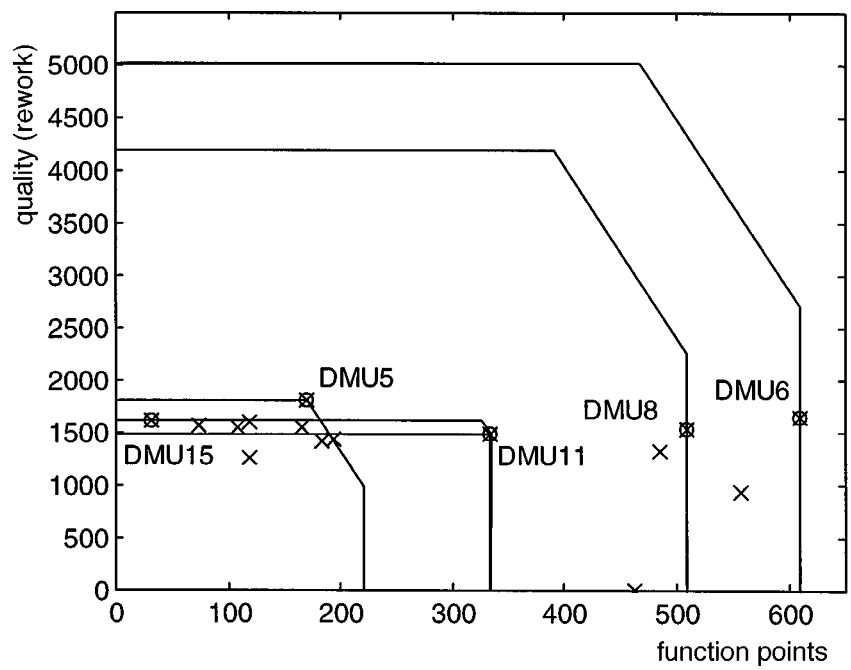

Figure 6b. Quality/functionality trade-offs – CRS.

the marginal rate of vector $\mathbf{v}$ to vector $\mathbf{u}$ as $\mathrm{MR}_{\mathbf{v},\mathbf{u}}(\mathbf{z}) \equiv \frac{\partial\alpha}{\partial\beta}\Big|_{\alpha,\beta=0}$. Thus, if the frontier is sufficiently smooth, one can show that this rate is given by

$$\mathrm{MR}_{\mathbf{v},\mathbf{u}}(\mathbf{z}) \equiv \left.\frac{\partial\alpha}{\partial\beta}\right|_{\alpha,\beta=0} = -\frac{|\partial F(\mathbf{x}+\beta\mathbf{u})/\partial\beta|_{\beta=0}}{|\partial F(\mathbf{x}+\alpha\mathbf{v})/\partial\alpha|_{\alpha=0}} \tag{6.2}$$

When the frontier is only piecewise differentiable, as in DEA, we can then define the marginal rates to the right and to the left as done in section 2.

Note that (6.2) reduces to the common marginal rates discussed in the previous sections, MR_{ij}, when $\mathbf{u} = \mathbf{e}_j$ and $\mathbf{v} = \mathbf{e}_i$; *i.e.* the unit vectors in the j-th and i-th directions, respectively. Then, it is straightforward to see that the computation of the more general marginal rates can be done using the techniques presented in Section 4 with the substitution of the unit vectors $\mathbf{e}_j$ and $\mathbf{e}_i$ by $\mathbf{u}$ and $\mathbf{v}$, respectively. Note also that the algorithm introduced in section 4.4 can now be used to generate general 2-D sections $\{(\alpha, \beta)\colon F(\mathbf{z}_o+\alpha\mathbf{v}+\beta\mathbf{u}) = 0\}$ for different $\mathbf{z}_o$.

Another common case arises when $\mathbf{u} = \frac{(-\mathbf{x},\mathbf{0})}{\|(-\mathbf{x},\mathbf{0})\|}$ and $\mathbf{v} = \frac{(\mathbf{0},\mathbf{y})}{\|(\mathbf{0},\mathbf{y})\|}$, the (unit) vectors that increase the inputs and outputs while remaining in the same proportions. With some simple manipulations, this leads to the formulation of returns to scale given in Golany and Yu (1994) and Hackman et al. (1994).

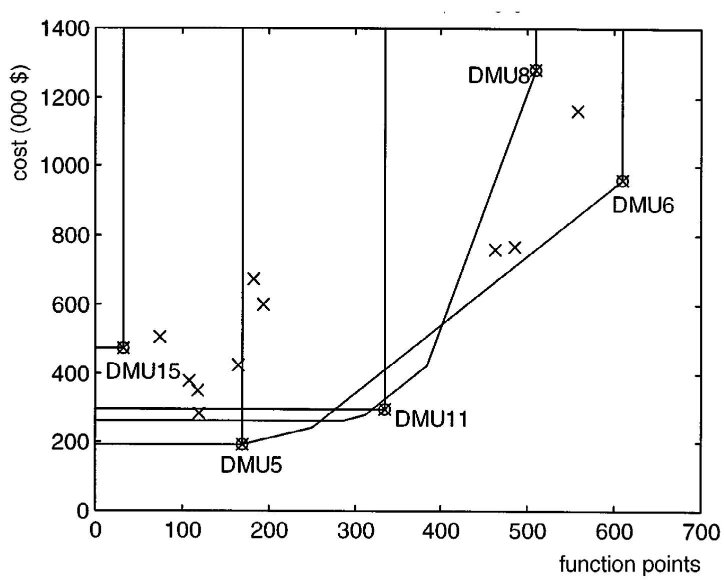

Figure 7. Cost/functionality trade-offs — VRS.

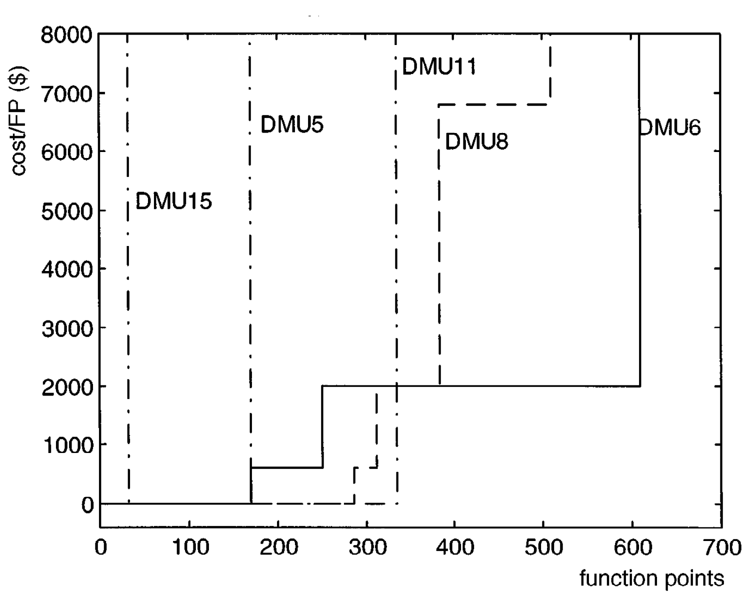

Figure 8. Cost/FP vs. FP — VRS.

Table A.1. Input-output data of simulated hospitals.

Hospital	Total cost ($ 000's)	Teaching Units	Regular Patients (000's)	Severe Patients (000's)
1	775.5	50	3	2
2	816.6	50	2	3
3	841.6	100	2	3
4	800.5	100	3	2
5	950.3	50	4	3
6	1,191.05	100	2	5
7	1,711.3	50	10	2
8	884.75	100	3	2
9	841.6	50	2	3
10	2,036.3	100	10	2
11	1,362.6	50	5	3
12	1,070	100	3	3
13	1,491.1	50	4	5
14	898.7	50	3	2
15	1,070	100	3	3

The analysis of general directional derivatives and general 2-D sections opens, in itself, an assortment of new possibilities and raises new interesting questions. From a practitioner's point of view, we may be able to analyze some simultaneous trade-offs within the production process. Clearly, the possibilities are endless. For example, a software project manager might be interested in maintaining a given proportion of quality and functionality while investigating the effect on costs for larger or smaller projects. Another interesting issue, with a more theoretical underpinning, is to analyze the sets of directions that might lead to either strictly positive or negative derivatives. While the marginal rates, as defined in Section 3, are always non-positive in DEA frontiers, the generalized rates (6.2) may take any value.

7. Concluding Remarks

This paper presents a general framework for analyzing marginal rates, and slopes in general, on DEA frontiers. The methods introduced allow the visualization as well as the practical and rigorous utilization of DEA multiplier information. Although the presentation was focused on the more commonly used BCC and CCR models, the concepts presented, and the algorithms derived from them, are quite general and can be readily applied to other mod-

Table A.2. Software projects: (transformed) data and efficiency scores.

Project	Cost $000's	FPs	Rework	Duration	θ (efficiency) CCR (constant RTS)	BCC (variable RTS)
1	1,162	557	931	0	0.42	0.72
2	766	485	1,328	0	0.56	0.86
3	378	108	1,552	0	0.44	0.51
4	421	165	1,555	211	0.61	0.62
5	194	169	1,813	96	1	1
6	961	609	1,651	39	0.56	1
7	673	183	1,420	185	0.34	0.37
8	1,279	509	1,527	217	0.35	1
9	284	119	1,599	190	0.87	0.88
10	507	74	1,558	240	0.55	0.63
11	294	334	1,484	267	1	1
12	348	118	1,254	158	0.58	0.66
13	598	194	1,428	115	0.33	0.35
14	759	462	0	102	0.54	0.8
15	471	32	1,609	255	0.62	1

els such as those with multiplier constraints, piecewise log-linear frontiers (multiplicative model), etc.

Piecewise linear DEA envelopment surfaces do not allow for unique derivatives at the extreme efficient points and, therefore, a useful characterization of these derivatives is given in terms of the ranges they can take at each given point. At a particular point, the bounds of these ranges are equivalent to the derivatives "to the right" and "to the left" of it. We present two computational methods to obtain these derivatives and show their connection to algorithms previously developed for the assessment of returns to scale. In particular, we provide a more general application of the modified simplex tableau algorithm introduced by Hackman et al. (1994) to generate all the rates at certain levels of the frontier (with the equivalent cost of two LP solutions) and to obtain graphical representations of *any* 2-D section of the frontier. Thus, by giving a complete picture of trade-offs and broadening the visual capabilities of the analysis, this set of tools should be of great practical importance to managerial applications. Indeed, it is our conviction that this kind of visual presentation of the results not only provides profound managerial information, but is vital for DEA to be adopted and used more readily by managers in industry (see also Schaffnit et al., 1995).

Another interesting potential application of this methodology is the construction and analysis of models with multiplier constraints. For example, complete knowledge of the trade-offs suggested exclusively by the input/output data (*i.e*. by a basic DEA model such as BCC) may be useful for estimating, *a priori*, the impact of subjective managerial goals on the results (efficiency scores, classification, etc.). Such information may also be helpful for adjusting, or even setting, the multiplier constraints themselves.

Acknowledgments

We wish to thank David Reese for useful comments on the manuscript and many fruitful discussions. This research was made possible by the generous financial support of SSHRC and NSERC of Canada, Toronto Dominion Bank, Royal Bank of Canada and Bell Canada.

Notes

1. More precisely, only the constraints for “extreme” efficient DMUs are necessary (*c.f.* Charnes et al., 1991).
2. A first order approximation means that, as h vanishes, the (absolute value of the) difference between the approximation and the derivative is less than, or equal to, a constant times h.
3. The treatment of a CCR frontier, when the section concerns one input and one output must be a little different, since the section will probably end with an unbounded ray. This can be handled without further complications.

References

Ali, A. I., and L. M. Seiford. (1993). “The Mathematical Programming Approach to Efficiency Analysis.” In H. O. Fried, K. C. A. Lovell, and S. S. Schmidt (eds.), *The Measurement of Productive Efficiency*. New York: Oxford University Press.

Ames, W. F. (1992). *Numerical Methods for Partial Differential Equations*, Third edition. New York: Academic Press Inc.

Banker, R. D., A. Charnes, and W. W. Cooper. (1984). “Some Models for Estimating Technical and Scale Efficiencies in Data Envelopment Analysis.” *Management Science* 30(9), 1078–1092.

Banker, R. D., and R. M. Thrall. (1992). “Estimation of Returns to Scale using Data Envelopment Analysis.” *European Journal of Operational Research* 62, 74–84.

Bessent, A., W. Bessent, J. Elam, and T. Clark. (1988). “Efficiency Frontier Determination by Constrained Facet Analysis.” *Operations Research* 36, 785–796.

Charnes, A., W. W. Cooper, B. Golany, L. Seiford, and J. Stutz. (1985). “Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions.” *Journal of Econometrics* 30, 91–107.

Charnes, A., W. W. Cooper, A. Lewin, and L. M. Seiford. (1995). *Data Envelopment Analysis: Theory, Methodology and Applications*. Kluwer Academic Publishers

Charnes, A., W. W. Cooper, and E. Rhodes. (1978). “Measuring the Efficiency of Decision Making Units.” *European Journal of Operational Research* 2(6), 429–444.

Charnes, A., W. W. Cooper, and R. M. Thrall. (1991). “A Structure for Classifying and Characterizing Efficiency and Inefficiency in Data Envelopment Analysis.” *Journal of Productivity Analysis* 2, 197–237.

Chilingerian, J. A. (1995). “Evaluating Physician Efficiency in Hospitals: A Multivariate Analysis of Best Practices.” *European Journal of Operational Research* 80, 548–574.

Dennis, Jr., J. E. (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood, New Jersey.

Fukuda, K. (1993). “cdd.c: C-Implementation of the Double Description Method for Computing All Vertices and Extremal Rays of a Convex Polyhedron given a System of Linear inequalities.” Dept. of Mathematics, Swiss Federal Inst. of Technology, Lausanne, Switzerland.

Golany, B., and G. Yu. (1994). "Estimating Returns to Scale in DEA." *European Journal of Operational Research*, forthcoming.

Hackman, S. T., U. Passy, and L. K. Platzman. (1994). "Explicit Representation of the Two-Dimensional Section of a Prodcution Possibility Set." *Journal of Productivity Analysis* 5, 161–170.

Paradi, J. C., D. N. Reese, and D. Rosen. (1995). "Applications of DEA to Measure the Efficiency of Software Production at Two Large Canadian Banks." *Annals of Operations Research*, in press.

Schaffnit, C., D. Rosen, and J. C. Paradi. (1995). "Envisioning DEA." Working paper, CMTE, University of Toronto, presented at the *INFORMS International Meeting*, Singapore, June 25–28, 1995.

Sherman, H. D. (1984). "Data Envelopment Analysis as a New Managerial Audit Methodology-Test and Evaluation." *Auditing: A Journal of Practice and Theory* 4, 33–35.

Thanassoulis, E. (1993). "A Comparison of Regression Analysis and Data Envelopment Analysis as Alternative Methods of Performance Assessment." *J. Operational Research Society* 44(11), 1129–1144.