



University of Connecticut

Department of Economics Working Paper Series

**Data Envelopment Analysis with
Alternative Returns to Scale**

by

Subhash C Ray
University of Connecticut

Working Paper 2018-20
November 2018

365 Fairfield Way, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
<http://www.econ.uconn.edu/>

This working paper is indexed in RePEc, <http://repec.org>

DATA ENVELOPMENT ANALYSIS WITH ALTERNATIVE RETURNS TO SCALE

Subhash C Ray

Department of Economics

University of Connecticut

Storrs CT 06269-1063

subhash.ray@uconn.edu

Abstract:

This paper offers an overview of Data Envelopment Analysis as a nonparametric method of measuring efficiency in production. Special attention is devoted to alternative returns to scale assumptions about the technology and identifying the local nature of returns to scale at projections of an inefficient unit on to the frontier. Both radial and non-radial measures of technical efficiency are considered.

JEL Classification: C61; D24

Keywords: Returns to Scale; Most Productive Scale Size; Radial and Non-Radial Efficiency;
Directional Distance Function

DATA ENVELOPMENT ANALYSIS WITH ALTERNATIVE RETURNS TO SCALE

Subhash C Ray

Department of Economics

University of Connecticut

Storrs CT 06269-1063

subhash.ray@uconn.edu

1. Introduction

Performance evaluation in any decision making situation involves comparing the outcome from the decision actually made with what is deemed to be the most preferred outcome within the constraints of the decision making problem. To evaluate the performance of a firm producing a single output from a specific bundle of inputs one compares its actual output with the maximum producible quantity from the bundle of inputs it is using. In textbook economics, the production function defines the maximum output producible from any given bundle of inputs. The actual output may fall below this maximum due to inefficiency. In practice, there is no readily available scientific formula showing the maximum output from a given input and the production function has to be constructed from observed input-output data.

The common practice in empirical research is to start with an explicit functional specification of the production function and to use regression to estimate the parameters of the model. However, the two-sided residuals imply that for some observations, the observed output actually exceeds the fitted value. This violates the assumption that the regression provides an upper limit on the producible output. In the Stochastic Frontier Analysis (SFA) literature this problem has been addressed in various ways. These include adjusting the intercept upwards to cover all data points from above (Greene 1980), specifying a one-sided distribution (like the Gamma distribution) for the error term (Richmond 1974), and a full blown stochastic frontier incorporating a one-sided error term representing inefficiency alongside a two-sided disturbance capturing random variation in the frontier (Aigner, Lovell, and Schmidt 1977).

An alternative to the parametric econometric analysis is the nonparametric approach of Data Envelopment Analysis (DEA) introduced by Charnes, Cooper, and Rhodes (1978). In neoclassical production theory the intellectual roots of DEA go all the way back to Debreu (1951), Koopmans (1953), Shephard (1953, 1970), Farrell (1957), Farrell and Fieldhouse (1962), Afriat (1972), and Hanoch and Rothschild (1972) among many others. In this strand of the efficiency literature, one makes only a number of general assumptions about the properties of the underlying production technology but leaves the production function unspecified. The observed input-output data are used to solve various mathematical

programming problems to measure the technical efficiency of a firm (as in Farrell (1957) or Farrell and Fieldhouse (1962) or to ascertain whether there exists any production technology satisfying the assumption relative to which the data would be consistent with optimizing behavior by the firms (Afriat (1972), Hanoch and Rothschild (1972), Diewert and Parkan (1983), Varian (1984), and Banker and Maindiratta (1988)).

The parametric approach of SFA has several advantages over DEA. Being an econometric approach it readily yields standard errors of the parameter estimates and allows application of standard statistical tests. Moreover, one can derive marginal productivities and various elasticities from the estimated model. However, validity of the entire empirical analysis rests critically on the validity of the functional form specified. If, for example, the true technology is log quadratic (Translog) but one estimates a Cobb Douglas production function, specification error may be interpreted as inefficiency. Moreover, the estimated model may violate regularity conditions like non-negative marginal productivities or negative own price elasticities of conditional input demand either globally or at individual data points. When the estimated model itself is invalid, little insight can be gained from standard errors of the coefficients.

In DEA the regularity conditions are imposed on the technology but no functional form is specified. Hence, (weak) monotonicity of the production function or convexity of isoquants are guaranteed to hold. A major weakness of DEA is that it treats all deviations from the frontier as inefficiency and does not readily accommodate random noise. It is possible, however, to generate an empirical distribution function of technical efficiency through bootstrapping and construct a confidence interval.

This chapter presents an overview of the DEA approach to measurement of technical efficiency. Special attention is paid to alternative returns to scale assumptions about the technology. The rest of the chapter is organized as follows. Section 2 introduces the production Possibility Set (PPS) as the foundation of neoclassical production economics and the Shephard Distance Function as a way to measure the proximity of an observed input output bundle to the frontier of the PPS. Section 3 details the nonparametric methodology of production analysis and measurement of output and input oriented radial technical efficiency under alternative returns to scale assumptions. Section 4 presents the so called ‘multiplier model’ form of the DEA linear programming problem and derives the CCR ratio measure of technical efficiency directly from the transformation function for a multiple output multiple input technology. Section 5 looks at the mathematical programming model of Aigner and Chu (1968) for estimating a deterministic parametric frontier production frontier and links it to DEA. Section 6 deals with measurement of scale efficiency and uses the Most Productive Scale Size (MPSS) to identify the nature of local returns to scale at different input-output bundles both those which are on the frontier and those which are below the frontier. Section 7 explains non-radial measures of technical efficiency. Section 8

covers Graph efficiency measurement including Graph Hyperbolic efficiency, Directional Distance Function, and Pareto-Koopmans efficiency. Section 9 is the conclusion.

2. Conceptual Foundations

The logical starting point of any discussion of production efficiency is the concept of the Production Possibility Set (PPS). An input vector $x^0 \in R_+^n$ and an output vector $y^0 \in R_+^m$ together constitute a feasible production plan if y^0 can be produced from x^0 . The production possibility set consists of all feasible production plans and is defined as

$$T = \{(x, y) : y \in R_+^m \text{ can be produced from } x \in R_+^n\}. \quad (1)$$

It is assumed that T is a closed set. In any given context the size and shape of the set T depends not only on the state of technical knowledge but also on a host of physical, legal, and cultural factors. In the single output case, it is a common practice to define a production function

$$y^* = f(x); \frac{\partial f}{\partial x_i} \geq 0, (i = 1, 2, \dots, n) \quad (2)$$

where y^* is the maximum quantity of output producible from a given input bundle x . The PPS can then be defined as

$$T = \{(x, y) : y \leq f(x), y \in R_+, x \in R_+\}. \quad (3)$$

Note that while y^* is the maximum output producible from the input bundle x , the actual output (y) can be less than y^* due to inefficiency. For multiple output multiple input production one uses the transformation function

$$F(x, y) = 0; \frac{\partial F}{\partial x_i} \leq 0, (i = 1, 2, \dots, n); \frac{\partial F}{\partial y_j} \geq 0, (j = 1, 2, \dots, m). \quad (4)$$

In that case

$$T = \{(x, y) : F(x, y) \leq 0; x \in R_+^n, y \in R_+^m\}. \quad (5)$$

The definition in (3) is a special case of the more general definition in (5) with $F(x, y) = y - f(x)$. If $F(x^0, y^0) = 0, (x^0, y^0)$ is a boundary point of T . Any reduction in all inputs or increase in all outputs

will result in a strictly positive value of the transformation function and the new input-output bundle will be infeasible. The *graph* of the technology¹ is the set

$$G = \{(x, y) : F(x, y) = 0; x \in R_+^m, x \in R_+^n\}. \quad (6)$$

If $F(x^0, y^0) < 0$, (x^0, y^0) is an interior point. Similarly, if $F(x^0, y^0) > 0$, (x^0, y^0) is infeasible.

2.1 Distance Functions and Technical Efficiency

Technical efficiency in production lies in producing the maximum output quantity (or a maximal output bundle) from a given bundle of inputs or in using a minimal input bundle to producing a given output quantity (or bundle). The question of efficiency in production was first formally addressed by Debreu (1951) who introduced the concept of the ‘Coefficient of Resource Utilization’. It was essentially a macroeconomic concept measuring production efficiency at the economy level. Technical efficiency of any individual producer depends on the location of its actual input-output bundle relative to the boundary of the PPS or the graph of the technology. Shephard (1953) introduced the Distance function for the one output production technology (subsequently generalized to the multiple output multiple input case by Shephard (1970) and McFadden (1978))² which provides a measure of technical efficiency of an individual production decision making unit.

The Output Distance Function evaluated at a specific input-output bundle (x^0, y^0) relative to the PPS (T) is

$$D^y(x^0, y^0) = \min \lambda : (x^0, \frac{1}{\lambda} y^0) \in T. \quad (6)$$

If one represents the PPS by the transformation function as in (5), the Output Distance Function becomes

$$D^y(x^0, y^0) = \min \lambda : F(x^0, \frac{1}{\lambda} y^0) \leq 0. \quad (7)$$

If $D^y(x^0, y^0) > 1$, the output bundle y^0 is not producible from x^0 so that $(x^0, y^0) \notin T$. On the other hand, if $D^y(x^0, y^0) \leq 1$, $F(x^0, y^0) \leq 0$ and $(x^0, y^0) \in T$. Of Course, if $D^y(x^0, y^0) = 1$, $(x^0, y^0) \in G$.

Thus, an alternative way to define the PPS is

$$T = \{(x, y) : D^y(x, y) \leq 1; x \in R_+^m, x \in R_+^n\}. \quad (8)$$

¹ Some writer prefer to call it the *Graph of the transformation function*.

² Although the volume was published in 1978, McFadden’s paper has been available since 1968.

Comparable to the Output Distance Function is the Input Distance Function

$$D^x(x^0, y^0) = \max \delta : F(\frac{1}{\delta} x^0, y^0) \leq 0. \quad (9)$$

Thus, yet another alternative definition of the PPS is

$$T = \{(x, y) : D^x(x, y) \geq 1; x \in R_+^m, x \in R_+^n\}. \quad (10)$$

As noted earlier, technical efficiency precludes the possibility of increasing outputs without increasing inputs or reducing inputs without reducing outputs. At this point, it is important to distinguish between *weak* and *strong* efficiency. An input-output bundle $(x, y) \in T$ is weakly efficient in its output orientation if $\alpha > 1 \Rightarrow (x, \alpha y) \notin T$. Similarly, $(x, y) \in T$ is weakly efficient in its input-orientation if

$\beta < 1 \Rightarrow (\beta x, y) \notin T$. As is quite apparent, weak efficiency rules out simultaneous increase in *all outputs* leaving the inputs unchanged or simultaneous reduction in *all inputs* without changing outputs. However, it does not rule out potential increase in one or several outputs or reduction in one or several inputs.³ By contrast, $(x, y) \in T$ is strongly output efficient if $y' \geq y \Rightarrow (x, y') \notin T$. Here the vector inequality $y' \geq y$ means that y' is at least as large as y in every coordinate and is strictly larger in some coordinate(s).

Thus, strong technical efficiency in output orientation rules out increasing any individual output without reducing any output or increasing any input. Similarly, $(x, y) \in T$ is strongly input efficient if

$x' \leq x \Rightarrow (x', y) \notin T$. Weak efficiency is also described as radial efficiency because it relates to equi-proportional change in all outputs or all inputs. Strong efficiency, on the other hand, is called non-radial efficiency.

The output-oriented radial technical efficiency of a firm producing output y^0 from the input x^0 is

$$\tau_y(x^0, y^0) = \frac{1}{\varphi^*} \quad (11)$$

where

$$\varphi^* = \max \varphi : (x^0, \varphi y^0) \in T. \quad (12)$$

Clearly,

$$\tau_y(x^0, y^0) = D^y(x^0, y^0). \quad (13)$$

³ This can happen if $\frac{\partial F}{\partial y_j} = 0$ for some outputs or $\frac{\partial F}{\partial x_i} = 0$ for some inputs.

The input-oriented radial technical efficiency is

$$\tau_x(x^0, y^0) = \min \theta : (\theta x^0, y^0) \in T. \quad (14)$$

This time,

$$\tau_x(x^0, y^0) = \frac{1}{D^x(x^0, y^0)}. \quad (15)$$

Farrell (1957) was the first to formally define and measure technical efficiency in production using input-output data.⁴ However, the first mathematical formulation of technical efficiency measurement as a Linear Programming (LP) optimization problem can be found in the appendix of Farrell and Fieldhouse (1962) where it is set up as an output oriented model.⁵

Measurement of output and input oriented technical efficiencies is shown graphically in Figures 1 and 2. In both figures the curve OP is the graph of the production function $y = f(x)$ and the points A and B are the input-output quantities (X_A, Y_A) and (X_B, Y_B) . In Figure 1, the points A^* and B^* are their output-oriented efficient projections. Hence, $\tau_y(X_A, Y_A) = \frac{AX_A}{A^*X_A}$ and $\tau_y(X_B, Y_B) = \frac{BX_B}{B^*X_B}$. Similarly, the points C and D are the input-oriented efficient projections of A and B . Thus, $\tau_x(X_A, Y_A) = \frac{CY_A}{AY_A}$ and $\tau_x(X_B, Y_B) = \frac{DY_B}{BY_B}$.

3. The Nonparametric methodology

In order to measure technical efficiency in any empirical application one has to define the production possibility set by specifying the production function (in the single output case) or the transformation function (in the multiple output case). In parametric models, this requires an explicit functional specification like the Cobb Douglas, Constant Elasticity of Substitution (CES), or the Translog production function. Measures of technical efficiency as well as elasticities of substitution between inputs derived from the calibrated model are all contingent upon the validity of the specified functional form. By contrast, in the nonparametric approach of Data Envelopment Analysis (DEA) one relies on a number of

⁴ The computation method in Farrell's paper was extremely burdensome. It involved comparing the input vector per unit of the output used by any unit with convex combinations of all other pairs of such (per unit) input bundles. The technical efficiency of the unit under evaluation was the minimum contraction factor feasible in its unit input bundle compared to any such convex combination. As is apparent from Diagram 2 on page 256 of Farrell (1957) it was an input oriented measure.

⁵ Of course, under the CRS assumption (as shown below) the output and input oriented measures of technical efficiency are identical.

quite general assumptions about the underlying technology and employs Mathematical Programming techniques to construct the frontier of the production possibility set.

Consider an industry producing m outputs from n inputs. Let (x^j, y^j) be the observed input-output bundle of firm j . Designate the sample of N observations as

$$\Omega = \{(x^j, y^j), j = 1, 2, \dots, N\} . \quad (16)$$

Assumptions

(A1) Every observed input-output bundle is feasible. Thus,

$$(x^j, y^j) \in \Omega \Rightarrow (x^j, y^j) \in T.$$

(A2) The production possibility set is convex.

$$(x^0, y^0) \in T \wedge (x^1, y^1) \in T \Rightarrow (\lambda x^0 + (1-\lambda)x^1, \lambda y^0 + (1-\lambda)y^1) \in T \forall \lambda \in (0,1).$$

(A3) Inputs are freely disposable.

$$(x^0, y^0) \in T \wedge x^1 \geq x^0 \Rightarrow (x^1, y^0) \in T.$$

(A4) Outputs are freely disposable.

$$(x^0, y^0) \in T \wedge y^1 \leq y^0 \Rightarrow (x^0, y^1) \in T.$$

One can use the data set Ω and the assumptions (A1-A4) to construct the production possibility set empirically as

$$\hat{T} = \left\{ (x, y) : x \geq \sum_{j=1}^N \lambda_j x^j; y \leq \sum_{j=1}^N \lambda_j y^j; \sum_{j=1}^N \lambda_j = 1; \lambda_j \geq 0; (j = 1, 2, \dots, N) \right\} \quad (17)$$

The intuition behind (17) is quite simple. By virtue of (A1) and (A2), every convex combination

$$(\bar{x}, \bar{y}) = \left(\sum_{j=1}^N \lambda_j x^j, \sum_{j=1}^N \lambda_j y^j \mid \sum_{j=1}^N \lambda_j = 1 \right) \text{ is feasible. Further, by (A3), } x \geq \bar{x} \text{ implies } (x, \bar{y}) \text{ is feasible.}$$

Finally by (A4) $y \leq \bar{y}$ implies (x, y) is feasible. The set \hat{T} is often described as the free disposal convex hull of the observed input-output bundles and is the smallest set satisfying assumptions (A1-A4). The frontier of this empirically constructed production possibility set *envelops* the observed data points most

tightly from above. Hence, measuring efficiency using this frontier as the benchmark for evaluation is called Data Envelopment Analysis (DEA).

3.1 Output and Input oriented Radial Technical Efficiency

Banker, Charnes, and Cooper (BCC) (1984) formulated the following LP model to measure the output oriented technical efficiency of a firm using input $x^0 \in R_+^n$ and producing output $y^0 \in R_+^m$:

$$\begin{aligned}
& \max \varphi \\
& s.t. \sum_{j=1}^N \lambda_j y_r^j \geq \varphi y_r^0 \quad (r = 1, 2, \dots, m); \\
& \quad \sum_{j=1}^N \lambda_j x_i^j \leq x_i^0 \quad (i = 1, 2, \dots, n); \\
& \quad \sum_{j=1}^N \lambda_j = 1; \\
& \quad \lambda_j \geq 0, (j = 1, 2, \dots, N); \varphi \text{ unrestricted.}
\end{aligned} \tag{18}$$

The optimal solution of (18) yields $\tau_y(x^0, y^0) = \frac{1}{\varphi}$.

It may be noted that even though φ is unrestricted, when (x^0, y^0) is one of the bundles in Ω (say the bundle of firm k), $(\lambda_k = 1, \lambda_j = 0 (j \neq k), \varphi = 1)$ is a feasible solution and, in that case, 1 would be a lower bound for φ . But even when (x^0, y^0) is not one of the observed bundles non-negativity of the λ s and the outputs will ensure that φ will never be negative⁶.

The benchmark input-output bundle for (x^0, y^0) is $\left(x^* = \sum_{j=1}^N \lambda_j^* x^j, y^* = \sum_{j=1}^N \lambda_j^* y^j \right)$ constructed from the optimal solution of the problem. For any output r the difference between the left hand side and the right hand side of the relevant output constraint, $s_r^+ = \sum_{j=1}^N \lambda_j^* y_r^j - \varphi^* y_r^0$, is the output slack representing the additional expansion of the output *feasible beyond the common expansion* by the scalar φ^* . Similarly, the input slack, $s_i^- = x_i^0 - \sum_{j=1}^N \lambda_j^* x_i^j$, is the potential reduction in input i . The scalar φ^* shows the factor by which *all outputs* can be expanded without requiring any additional input. In fact, some outputs can be expanded

⁶ However, if any individual input in the bundle x^0 is smaller than the smallest value of the corresponding input across all observations in the data set Ω , (14) will not have a feasible solution.

further when there are positive output slacks. Similarly, some inputs can even be reduced if there are positive input slacks. The BCC output oriented model yields a *radial* measure of technical efficiency because it is the inverse of the radial output expansion factor φ^* and does not incorporate the output slacks.

In the single output case the optimal value of the objective function in the output-oriented DEA problem (φ^*) yields an *estimate* of the maximum output producible from the input bundle x^0 as

$$\hat{f}(x^0) = \varphi^* y_0. \quad (19)$$

The true maximum output that can be produced from x^0 may actually be considerably higher than $\varphi^* y_0$. But we cannot infer that on the basis of the observed input-output bundles without making additional assumptions about the technology. However, *it cannot be any smaller than* $\varphi^* y_0$ if the assumptions (A1-A4) hold. In that sense, it is the most conservative estimate of the frontier output and, hence, $\tau_y(x^0, y_0) = \frac{1}{\varphi^*}$ is an upper bound on the output orient technical efficiency of the firm.⁷

The corresponding input oriented technical efficiency of the firm using input x^0 and producing output y^0 may be evaluated as

$$\tau_x(x^0, y^0) = \theta^*, \text{ where}$$

$$\begin{aligned} \theta^* &= \min \theta \\ \text{s.t. } &\sum_{j=1}^N \lambda_j y_r^j \geq y_r^0 \quad (r = 1, 2, \dots, m); \\ &\sum_{j=1}^N \lambda_j x_i^j \leq \theta x_i^0 \quad (i = 1, 2, \dots, n); \\ &\sum_{j=1}^N \lambda_j = 1; \\ &\lambda_j \geq 0, (j = 1, 2, \dots, N); \theta \text{ unrestricted.} \end{aligned} \quad (20)$$

Again, it is obvious that $0 < \theta^* \leq 1$. It should be noted that the benchmark input-output bundle

$\left(x^* = \sum_{j=1}^N \lambda_j^* x^j, y^* = \sum_{j=1}^N \lambda_j^* y^j \right)$ on the frontier for the input oriented DEA problem, will, in general, be different from what was obtained for the output oriented problem.

⁷ We use superscripts for vectors and subscripts for scalars.

3.2 Constant Returns to Scale

If one assumes constant returns to scale (CRS), we get the additional assumption

$$(A5) \ (x, y) \in T \Rightarrow (kx, ky) \in T \ \forall k \geq 0.$$

An implication of the CRS assumption is that in the single output case the production function $y^* = f(x)$ is homogenous of degree 1. That is, $f(kx) = kf(x)$. In the multiple output case, CRS implies that the transformation function is homogeneous of degree 0. That is $F(kx, ky) = F(x, y)$. This ensures that if $F(x^0, y^0) \leq 0$, then $F(kx^0, ky^0) \leq 0$. Hence, if (x^0, y^0) is feasible, so is (kx^0, ky^0) .

Under the additional assumption of CRS, the empirically constructed PPS is

$$\hat{T}_C = \left\{ (x, y) : x \geq \sum_{j=1}^N \lambda_j x^j; y \leq \sum_{j=1}^N \lambda_j y^j; \lambda_j \geq 0; (j = 1, 2, \dots, N) \right\} \quad (21)$$

To understand why the constraint $\sum_{j=1}^N \lambda_j = 1$ is no longer included consider the following. We have seen

above that under (A1-A2), $(\bar{x}, \bar{y}) = \left(\sum_{j=1}^N \lambda_j x^j, \sum_{j=1}^N \lambda_j y^j \mid \sum_{j=1}^N \lambda_j = 1 \right)$ is feasible. But now, with the added assumption of CRS, $(k\bar{x}, k\bar{y}) = \left(\sum_{j=1}^N \lambda_j x^j, \sum_{j=1}^N \lambda_j y^j \mid \sum_{j=1}^N \lambda_j = k; k \geq 0 \right)$ is feasible for any $k \geq 0$. But non-

negativity of k is automatically satisfied by the non-negativity constraints on the λ_j s and no additional constraint on the sum of the λ s is needed.

The CCR output oriented CRS DEA LP model is

$$\begin{aligned} \varphi_C^* &= \max \varphi_C \\ s.t. \sum_{j=1}^N \lambda_j y_r^j &\geq \varphi_C y_r^0 \ (r = 1, 2, \dots, m); \\ \sum_{j=1}^N \lambda_j x_i^j &\leq x_i^0 \ (i = 1, 2, \dots, n); \\ \lambda_j &\geq 0, (j = 1, 2, \dots, N); \varphi \text{ unrestricted.} \end{aligned} \quad (22)$$

The CRS output oriented radial technical efficiency is

$$\tau_y^C(x^0, y^0) = \frac{1}{\varphi_C^*}. \quad (23)$$

The corresponding CRS input oriented model is

$$\begin{aligned}
& \theta_C^* = \min \theta \\
& s.t. \sum_{j=1}^N \lambda_j y_r^j \geq y_r^0 \quad (r = 1, 2, \dots, m); \\
& \sum_{j=1}^N \lambda_j x_i^j \leq \theta x_i^0 \quad (i = 1, 2, \dots, n); \\
& \lambda_j \geq 0, (j = 1, 2, \dots, N); \theta \text{ unrestricted.}
\end{aligned} \tag{24}$$

The CRS input oriented technical efficiency is

$$\tau_x^C(x^0, y^0) = \theta_C^*. \tag{25}$$

To prove that under CRS the input and output oriented radial measures of technical efficiency are identical we first rewrite the objective function in (22) as $\psi_C^* = \min \frac{1}{\varphi_C}$ and then divide the constraints by φ_C to rewrite the problem as

$$\begin{aligned}
& \min \frac{1}{\varphi_C} \\
& s.t. \sum_{j=1}^N \frac{\lambda_j}{\varphi_C} y_r^j \geq y_r^0 \quad (r = 1, 2, \dots, m); \\
& \sum_{j=1}^N \frac{\lambda_j}{\varphi_C} x_i^j \leq \frac{1}{\varphi_C} x_i^0 \quad (i = 1, 2, \dots, n); \\
& \lambda_j \geq 0, (j = 1, 2, \dots, N); \varphi_C \text{ unrestricted.}
\end{aligned} \tag{26}$$

Now define $\psi = \frac{1}{\varphi_C}$ and $\mu_j = \frac{\lambda_j}{\varphi_C}$. Note that, as explained before, even though in principle φ_C is unrestricted in sign, non-negativity of outputs ensures that it will never be negative. Hence μ_j will also be non-negative for each observation j . Hence, the CRS output oriented problem can be reduced to

$$\begin{aligned}
& \min \psi \\
& s.t. \sum_{j=1}^N \mu_j y_r^j \geq y_r^0 \quad (r = 1, 2, \dots, m); \\
& \sum_{j=1}^N \mu_j x_i^j \leq \psi x_i^0 \quad (i = 1, 2, \dots, n); \\
& \mu_j \geq 0, (j = 1, 2, \dots, N); \psi \text{ unrestricted.}
\end{aligned} \tag{27}$$

The problem in (27) is exactly the same as the input oriented problem in (24). Hence, ψ^* in (27) equals θ_C^* in (24) and $\frac{1}{\phi_C^*}$ from (22). This proves that $\tau_y^C(x^0, y^0) = \tau_x^C(x^0, y^0)$.

4. The CCR Ratio and the Multiplier Model

In their seminal paper introducing DEA Charnes, Cooper, and Rhodes (1978) defined technical efficiency as

$$\begin{aligned}
 h(x^0, y^0) = \max \quad & \frac{\sum_{r=1}^m u_r y_r^0}{\sum_{i=1}^n v_i x_i^0} \\
 \text{s.t.} \quad & \frac{\sum_{r=1}^m u_r y_r^j}{\sum_{i=1}^n v_i x_i^j} \leq 1; (j = 1, 2, \dots, N) \\
 & u_r, v_i \geq 0; (r = 1, 2, \dots, m; i = 1, 2, \dots, n).
 \end{aligned} \tag{28}$$

In the OR/MS literature, the numerator $\sum_{r=1}^m u_r y_r^0$ is described as the *virtual output* and the denominator

$\sum_{i=1}^n v_i x_i^0$ as the *virtual input* of the unit under evaluation. In that sense, $h(x^0, y^0)$ is a measure of total

factor productivity rather than of technical efficiency. Because $(x^0, y^0) \in \Omega$ the constraint in (28) ensures that $h(x^0, y^0)$ cannot exceed unity. Further, because the objective function and the constraints

are all homogeneous of degree 0 in u and v , we can normalize the aggregation weights as $\sum_{i=1}^n v_i x_i^0 = 1$.

The linear fractional functional programming problem in (28) can then be equivalently written as the following LP problem:

$$\begin{aligned}
h(x^0, y^0) &= \max \sum_{r=1}^m u_r y_r^0 \\
s.t. \quad & \sum_{r=1}^m u_r y_r^j - \sum_{i=1}^n v_i x_i^j \leq 0; (j = 1, 2, \dots, N) \\
& \sum_{i=1}^n v_i x_i^0 = 1; \\
& u_r, v_i \geq 0; (r = 1, 2, \dots, m; i = 1, 2, \dots, n).
\end{aligned} \tag{29}$$

The problem in (29) is the linear programming dual of the optimization problem for the CRS input oriented technical efficiency in (24). Hence, by standard duality results, the CCR Ratio $h(x^0, y^0)$ in (29) equals $\tau_x^C(x^0, y^0)$ from (24) and under CRS also equals $\tau_y^C(x^0, y^0)$. Ray (2017) provides a derivation of the CCR ratio directly from the transformation function. Consider the input-output bundle (x^0, y^0) and assume that $F(x^0, y^0) < 0$ so that (x^0, y^0) is an inefficient bundle. Next, consider the Shephard Distance Function

$$D^y(x^0, y^0) = \min \beta : (x^0, \frac{1}{\beta} y^0) \in T \Leftrightarrow F(x^0, \frac{1}{\beta} y^0) = 0. \tag{30}$$

It is the inverse of the largest scalar δ such that $F(x^0, \delta y^0) = 0$ and is the same as the output oriented Farrell efficiency. Clearly, for (x^0, y^0) , $\delta > 1$ and $\beta < 1$.

Focus now on the efficient input-output bundle $(x^0, y_0^*) = (x^0, \delta y^0)$ lying on the graph of the technology. Thus $F(x^0, y_0^*) = 0$. Now, due to homogeneity of degree 0,

$$F(x^0, y_0^*) = \sum_i \left(\frac{\partial F}{\partial x_i} \right)_{x^0, y_0^*} x_{i0} + \sum_r \left(\frac{\partial F}{\partial y_r} \right)_{x^0, y_0^*} y_{r0}^* = 0. \tag{31}$$

$$\text{Define } F_i^0 \equiv \left(\frac{\partial F}{\partial x_i} \right)_{x^0, y_0^*} (i = 1, 2, \dots, n) \tag{32}$$

$$\text{and } F_r^0 \equiv \left(\frac{\partial F}{\partial y_r} \right)_{x^0, y_0^*} (r = 1, 2, \dots, m). \tag{33}$$

Then (31) becomes

$$\sum_i F_i^0 x_{i0} + \sum_r F_r^0 y_{r0}^* = 0. \tag{34}$$

Because $y_{r0}^* = \delta y_{r0}$, (34) leads to

$$\delta \sum_r F_r^0 y_{r0} = - \sum_i F_i^0 x_{i0}. \quad (35)$$

Thus,

$$\beta = \frac{1}{\delta} = \frac{\sum_r F_r^0 y_{r0}}{- \sum_i F_i^0 x_{i0}}. \quad (36)$$

Finally, define $u_r \equiv F_r^0$ and $v_i \equiv -F_i^0$ to derive

$$\beta = \frac{1}{\delta} = \frac{\sum_r u_r y_{r0}}{\sum_i v_i x_{i0}}. \quad (37)$$

This shows that the CCR ratio in (28) is the same as the Shephard Distance Function or the Farrell

measure of technical efficiency. Finally, $\sum_{r=1}^m u_r y_r - \sum_{i=1}^n v_i x_i = 0$ is a supporting hyperplane to the PPS at

$(x^0, \delta^* y^0)$ and due to convexity $\sum_{r=1}^m u_r y_r^j - \sum_{i=1}^n v_i x_i^j \leq 0 \forall (x^j, y^j) \in \Omega$. The multipliers (u and v) are the vectors of shadow prices of outputs and inputs. In that sense, the CCR ratio is the ‘shadow return on outlay’ and the inequality constraints are comparable to the ‘no positive profit’ condition in a competitive market under CRS.

The dual LP problem for the output oriented VRS DEA (also known as the BCC) model is

$$\begin{aligned} \min \quad & v_0 + \sum_{i=1}^n v_i x_i^0 \\ \text{s.t.} \quad & v_0 + \sum_{i=1}^n v_i x_i^j - \sum_{r=1}^m u_r y_r^j \geq 0; (j = 1, 2, \dots, N) \\ & \sum_{i=1}^n u_r y_r^0 = 1; \\ & u_r, v_i \geq 0; (r = 1, 2, \dots, m; i = 1, 2, \dots, n). \end{aligned} \quad (38)$$

In the next section, we show how the BCC DEA problem relates to the deterministic parametric frontier model of Aigner and Chu (1968).

5. A Deterministic Parametric Frontier as a special DEA Problem

As noted earlier, in the single output case the observed input-output data must satisfy the inequality $y \leq f(x)$. It is convenient, therefore to write the deterministic production frontier⁸ as

$$y = f(x).e^{-u}, u \geq 0 \quad (39)$$

Aigner and Chu (1968) specified the Cobb Douglas form of the frontier production function

$$f(x) = Ax_1^{\beta_1} x_2^{\beta_2} \dots x_k^{\beta_k}. \quad (40)$$

Then (39) can be written in logarithmic form for each individual observation j as

$$u_j = \beta_0 + \beta_1 \ln x_1^j + \beta_2 \ln x_2^j + \dots + \beta_k \ln x_k^j - \ln y_j \geq 0 \quad (41)$$

Given the one-sided distribution of u , the usual OLS estimation procedure does not work in this case.

Aigner and Chu proposed minimizing either $\sum_{j=1}^N u_j^2$ or $\sum_{j=1}^N u_j$ subject to the constraints that each estimated

residual u_j is non-negative. Note that minimizing $\sum_{j=1}^N u_j$ is equivalent to minimizing

$$\frac{1}{N} \sum_{j=1}^N u_j = \bar{u} = \beta_0 + \beta_1 \overline{\ln x_1} + \beta_2 \overline{\ln x_2} + \dots + \beta_k \overline{\ln x_k} - \overline{\ln y} \quad (42)$$

Further, $\overline{\ln y}$ is a constant and plays no role in the minimization problem. Hence, the Aigner-Chu problem can be formulated as

$$\begin{aligned} \min & \beta_0 + \beta_1 \overline{\ln x_1} + \beta_2 \overline{\ln x_2} + \dots + \beta_k \overline{\ln x_k} \\ \text{s.t.} & \beta_0 + \beta_1 \ln x_1^j + \beta_2 \ln x_2^j + \dots + \beta_k \ln x_k^j \geq \ln y_j \\ & \beta_i \geq 0 (i = 1, 2, \dots, k); \beta_0 \text{ unrestricted.} \end{aligned} \quad (43)$$

The LP dual to (43) is

⁸ In the parametric frontier production function literature, (39) is described as a deterministic frontier. It needs to be made clear that what is deterministic above is the frontier output $y^* = f(x)$ but the actual output y is stochastic because e^{-u} , the inefficiency component, is still random.

$$\begin{aligned}
& \max \sum_{j=1}^N \lambda_j \ln y_j \\
& s.t. \sum_{j=1}^N \lambda_j \ln x_i^j \leq \overline{\ln x_i} \quad (i = 1, 2, \dots, m) \\
& \sum_{j=1}^N \lambda_j = 1; \lambda_j \geq 0 \quad (j = 1, 2, \dots, N)
\end{aligned} \tag{44}$$

Next, define $\varphi \equiv \frac{\sum_{j=1}^N \lambda_j \ln y_j}{\ln y}$. Because $\lambda_j = \frac{1}{N}$ for all j is a feasible solution for (44),

$\sum_{j=1}^N \lambda_j \ln y_j \geq \varphi \overline{\ln y}$. Hence, (44) can be equivalently written as

$$\begin{aligned}
& \max \varphi \\
& s.t. \sum_{j=1}^N \lambda_j \ln y_j \geq \varphi \overline{\ln y}; \\
& \sum_{j=1}^N \lambda_j \ln x_i^j \leq \overline{\ln x_i} \quad (i = 1, 2, \dots, m) \\
& \sum_{j=1}^N \lambda_j = 1; \lambda_j \geq 0 \quad (j = 1, 2, \dots, N)
\end{aligned} \tag{45}$$

Thus, the Aigner-Chu model actually solves the BCC DEA output oriented problem evaluating the efficiency of the geometric mean of the input-output bundle in the (log) input-output space.⁹

6. Scale Efficiency

The concept of scale efficiency is best explained for a 1-output 1-input technology. Consider a firm with the input-output pair (x_0, y_0) . At the observed point, the average productivity of the firm is $AP_0 = \frac{y_0}{x_0}$.

If the firm is inefficient, its output can be increased to $f(x_0) > y_0$ and as result its average productivity would increase to $AP(x_0) = \frac{f(x_0)}{x_0}$. Thus, eliminating technical inefficiency would automatically raise

average productivity. However, once the firm has been projected on to the frontier, no further increase in average productivity is possible without changing the input level. All points on the frontier are technically efficient. However, unless CRS holds globally across the frontier, average productivity will vary across different points on the frontier. We may define average productivity along the frontier as $AP(x) = \frac{f(x)}{x}$.

⁹ For an MLE interpretation of the Aigner-Chu model, see Schmidt (1976).

Frisch (1965) defined the input level, x^* , where $AP(x)$ reaches a maximum as the *technically optimal production scale* (TOPS). A measure of scale efficiency of the firm would then be

$$SE(x_0) = \frac{AP(x_0)}{AP(x^*)}. \quad (46)$$

It is well known from standard microeconomics that the first order condition for a maximum of $AP(x)$ is $\frac{f(x)}{x} = f'(x) \Leftrightarrow f(x) = xf'(x)$. Hence, at the input level x^* , $AP(x^*) = f'(x^*)$. Define the constant $\kappa = f'(x^*)$ and a counterfactual CRS production function

$$r(x) = \kappa x. \quad (47)$$

Two things may be emphasized. First, by the definition of the TOPS, for every input level x

$$\begin{aligned} AP(x) &= \frac{f(x)}{x} \leq AP(x^*) = \frac{f(x^*)}{x^*} = f'(x^*) = \kappa \\ \Rightarrow f(x) &\leq \kappa x = r(x). \end{aligned} \quad (48)$$

Second,

$$SE(x_0) = \frac{AP(x_0)}{AP(x^*)} = \frac{f(x_0)}{\kappa x_0} = \frac{f(x_0)}{r(x_0)} \leq 1. \quad (49)$$

One can also write (49) as

$$SE(x_0) = \frac{f(x_0)}{r(x_0)} = \frac{\frac{y_0}{r(x_0)}}{\frac{y_0}{f(x_0)}}. \quad (50)$$

Now, $y^* = f(x)$ is the true VRS frontier and $\frac{y_0}{f(x_0)} = \tau_y(x_0, y_0)$ is a measure of the output oriented VRS technical efficiency. On the other hand, $y^{**} = r(x)$ is an artificial CRS frontier and $\frac{y_0}{r(x_0)} = \tau_y^C(x_0, y_0)$ is a measure of the output oriented CRS technical efficiency. Hence, a simple measure of scale efficiency is

$$SE(x_0) = \frac{\tau_y^C(x_0, y_0)}{\tau_y(x_0, y_0)}. \quad (51)$$

. In Figure 3 the point A shows the input-output bundle (x_0, y_0) and the point B on the production function $y^* = f(x)$ is the output oriented technically efficient projection. The most productive input scale is x^* and $AP(x^*) = \frac{f(x^*)}{x^*} = \frac{Cx^*}{Ox^*} = \frac{Dx_0}{Ox_0}$. Also, the tangent to the production function at the point C can be treated as a counterfactual CRS production function $y^{**} = r(x) = \kappa x; \kappa \equiv f'(x^*)$. Thus,

$$SE(x_0) = \frac{\frac{f(x_0)}{x_0}}{\frac{f(x^*)}{x^*}} = \frac{Bx_0}{Dx_0} = \frac{\frac{Ax_0}{Dx_0}}{\frac{Ax_0}{Bx_0}} = \frac{f(x_0)}{r(x_0)} = \frac{\frac{y_0}{r(x_0)}}{\frac{y_0}{f(x_0)}} = \frac{D_C^y(x_0, y_0)}{D^y(x_0, y_0)}.$$

6.1 Ray Average Productivity and Returns to Scale

The concept of average productivity is unequivocal only in a single-output single-input case. When multiple inputs are used for production (as is almost universally the case in real life) one can measure partial average productivities but to obtain a single measure of total factor productivity one must aggregate the individual inputs into a scalar. In the multiple-output case, the individual outputs also need to be aggregated. One way to avoid such aggregation is to consider only proportional variations in all inputs and the consequent proportional variation in all outputs. Consider a technically efficient pair of input-output bundles (x^0, y^0) satisfying $F(x^0, y^0) = 0$. Now consider another input bundle $x^1 = \beta x^0$. Thus the two input bundles have the same input-proportions and differ only in scale. Next consider an output bundle y satisfying $F(x^1, y) = 0$. There will be many output bundles y satisfying $F(x^1, y) = 0$. Out of them we select the one which is proportional to y^0 . Denote this as $y^1 = \alpha y^0$. Now compare the bundles (x^0, y^0) and $(x^1, y^1) = (\beta x^0, \alpha y^0)$. Note that because outputs increase (decrease) with inputs, $\beta > 1$ implies $\alpha > 1$. If we treat x^0 as a single unit of a composite input and y^0 as a single unit of a composite output then (x^1, y^1) corresponds to β units of the composite input and α units of the composite output. In terms of the composite input and output, the *ray average productivity* at (x^0, y^0) is 1 and at (x^1, y^1) it is $\frac{\alpha}{\beta}$. Now, if $\alpha > \beta$, we conclude that ray average productivity has increased and increasing returns to scale holds locally at (x^0, y^0) . Similarly, locally DRS holds when $\frac{\alpha}{\beta} < 1$. Of particular interest is a bundle for which $\frac{\alpha}{\beta} = 1$ showing locally CRS.

Note that between any two points (x, y) and $(x + dx, y + dy)$ both on the frontier

$$\begin{aligned} dF &= \sum_{i=1}^n F_i dx_i + \sum_{j=1}^m F_j dy_j = 0 \\ \Rightarrow \sum_{i=1}^n F_i x_i \frac{dx_i}{x_i} + \sum_{j=1}^m F_j y_j \frac{dy_j}{y_j} &= 0 \end{aligned} \tag{52}$$

But when the input bundles are proportional, $\frac{dx_i}{x_i}$ equals some constant q_1 for every input i . Similarly, proportionality of the output bundles implies $\frac{dy_j}{y_j}$ equals a constant q_2 for every output j . Hence, from (52)

$$\begin{aligned}
q_1 \sum_{i=1}^n F_i x_i + q_2 \sum_{j=1}^m F_j y_j &= 0 \\
\Rightarrow \frac{q_2}{q_1} &= \frac{-\sum_{i=1}^n F_i x_i}{\sum_{j=1}^m F_j y_j}.
\end{aligned} \tag{53}$$

Now, $x + dx = (1 + q_1)x$ and $y + dy = (1 + q_2)y$. Define $\beta = 1 + q_1$ and $\alpha = 1 + q_2$. Then $q_2 > q_1$ implies $\alpha > \beta$ and locally IRS holds at (x, y) . Similarly, $q_2 < q_1$ implies locally DRS. Finally, $q_2 = q_1$ implies locally CRS. Starrett (1977) defined the *degree of increasing returns* as

$$DIR = \frac{q_2}{q_1} - 1. \tag{54}$$

6.2 Most Productive Scale Size and Local Returns to Scale

Banker (1984) defined the *most productive scale size* (MPSS) in the context of multiple output multiple input production as a generalization of Frisch's concept of the TOPS. According to his definition,

an input-output bundle (x^0, y^0) is an MPSS if for all non-negative real numbers (α, β) such that $(\beta x^0, \alpha y^0)$ is feasible, $\frac{\alpha}{\beta} \leq 1$.

In other words, (x^0, y^0) is an MPSS if and only if there is no other input output bundle proportional to it with a higher ray average productivity. Obviously, an inefficient bundle cannot be an MPSS because both its output oriented projection $(x^0, \phi^* y^0)$ and its input oriented projection $(\theta^* x^0, y^0)$ will have a higher ray average productivity.

6.3 Identifying Local Returns to Scale at an Efficient Bundle

In the DEA literature there are three alternative ways to identify the local returns to scale properties of an input-output bundle that lies on the frontier of the PPS (a) Banker's Primal approach, (b) a Dual approach to due to BCC, and (c) a Nesting approach due to Färe, Grosskopf, and Lovell (FGL) (1985). Of course, all of them lead to the same conclusion.

Banker's Primal Approach

Assume that the bundle (x^0, y^0) is efficient under VRS. The following theorem in Banker (1984) identifies whether or not it is an MPSS.

Theorem: An input-output bundle (x^0, y^0) is an MPSS if and only if the optimal value of the objective function of a CCR-DEA model equals unity for this input-output combination.

Proof: See Banker (1984, p 40).

This theorem only determines if the bundle (x^0, y^0) is an MPSS. If it is not, all we know is that locally CRS does not hold at this point on the frontier. That does not, by itself, pick between IRS and DRS. However, the following corollaries from the theorem can be used to identify local RTS properties by looking at the optimal solution of the CCR DEA problem:

1. If $k = \sum_{j=1}^N \lambda_j^* = 1$, (x^0, y^0) is an MPSS and CRS holds locally.
2. If $k = \sum_{j=1}^N \lambda_j^* < 1$, IRS holds locally at (x^0, y^0) .
3. If $k = \sum_{j=1}^N \lambda_j^* > 1$, DRS holds locally at (x^0, y^0) .

Note that the corollaries (1-3) hold both for the output-oriented problem in (22) and the input-oriented problem in (24). One potential problem with this approach is that there may be alternative optimal solutions to the CCR DEA LP problem with the optimal values of the λ s adding up to greater than 1 in one and less than 1 in another. To resolve this ambiguity, one should modify the corollaries above and require that the respective conditions for IRS and DRS should hold at all optimal solutions of the CCR DEA problem.

To implement this, one would first solve either the output or the input oriented CRS DEA problem (22) or (24) and obtain the optimal value. Suppose that one selected the output oriented model and obtained ϕ^* from (22). Next one would use ϕ^* as a parameter to solve the following problem

$$\begin{aligned}
 \max k &= \sum_{j=1}^N \lambda_j \\
 \text{s.t. } \sum_{j=1}^N \lambda_j y_r^j &\geq \phi^* y_r^0 \quad (r = 1, 2, \dots, m); \\
 \sum_{j=1}^N \lambda_j x_i^j &\leq x_i^0 \quad (i = 1, 2, \dots, n); \\
 \lambda_j &\geq 0, (j = 1, 2, \dots, N).
 \end{aligned} \tag{55}$$

One would conclude that the condition in corollary 2 holds in all optimal solutions of (22) if the optimal k^* in (55) is less than 1.

To check for the condition in corollary 3, one would set up (55) as a minimization problem and conclude that DRS holds at (x^0, y^0) if the minimum of k is greater than 1.

The BCC Dual Approach

Banker, Charnes, and Cooper (BCC) (1984) offer an alternative method of identifying local returns to scale from the following dual of the output-oriented VRS DEA problem shown in (38) above.

BCC have shown that

- (i) CRS holds at (x^0, y^0) if at the optimal solution of (38) v_0 is zero;
- (ii) IRS holds at (x^0, y^0) if at the optimal solution of (38) v_0 is < 0 ;
- (iii) DRS holds at (x^0, y^0) if at the optimal solution of (38) v_0 is > 0 .

A simple proof of (i)-(iii) follows from Ray (2017). Because (x^0, y^0) is efficient by assumption, at the optimal solution of (38)

$$\begin{aligned} v_0^* + \sum_{i=1}^n v_i^* x_i^0 &= 1 = \sum_{r=1}^m u_r^* y_r^0 \\ \Rightarrow v_0^* + \sum_{i=1}^n v_i^* x_i^0 - \sum_{r=1}^m u_r^* y_r^0 &= 0. \end{aligned} \tag{56}$$

But, as defined above $u_r \equiv F_r^0$ and $v_i \equiv -F_i^0$ and in the present context (53) can also be written as

$$\begin{aligned} \frac{q_2}{q_1} &= \frac{-\sum_{i=1}^n F_i x_i}{\sum_{j=1}^m F_j y_j} = \frac{\sum_{i=1}^n v_i^* x_i^0}{\sum_{r=1}^m u_r^* y_r^0} \\ \Rightarrow \frac{q_2}{q_1} - 1 &= \frac{\sum_{i=1}^n v_i^* x_i^0 - \sum_{r=1}^m u_r^* y_r^0}{\sum_{r=1}^m u_r^* y_r^0}. \end{aligned} \tag{57}$$

Because the denominator in (57) is always positive, the sign of the ratio is determined by the sign of the numerator. Specifically, when IRS holds,

$$\begin{aligned} \frac{q_2}{q_1} - 1 &> 0 \\ \Rightarrow -\sum_{i=1}^n v_i^* x_i^0 - \sum_{j=1}^m u_j^* y_j^0 &= -v_0^* > 0 \quad (58) \\ \Rightarrow v_0^* &< 0. \end{aligned}$$

By the same logic, DRS implies $\frac{q_2}{q_1} - 1 > 0$ and $v_0^* > 0$. Finally, for CRS $q_1 = q_2$ and $v_0^* = 0$. Of course, as in the case of Banker's approach, multiple optimal solutions pose a problem and the conditions (ii) and (iii) have to be appropriately modified.

A Nesting Approach

Färe, Grosskopf, and Lovell (FGL) (1985) consider a technology that lies in between CRS and the VRS technologies. They call it a non-increasing returns to scale (NIRS) technology. Under the assumption of NIRS

$$(x^0, y^0) \in T \Rightarrow (kx^0, ky^0) \in T; 0 \leq k \leq 1. \quad (59)$$

Thus, any feasible input-output bundle remains feasible if it is scaled downwards but not necessarily feasible if scaled upwards.

The DEA estimate of an NIRS production possibility set is

$$\hat{T}^N = \left\{ (x, y) : x \geq \sum_{j=1}^N \lambda_j x^j; y \leq \sum_{j=1}^N \lambda_j y^j; \sum_{j=1}^N \lambda_j \leq 1, \lambda_j \geq 0; (j = 1, 2, \dots, N) \right\} \quad (60)$$

Note that the three different sets in (17), (21), and (60) are nested so that $\hat{T} \subset \hat{T}^N \subset \hat{T}^C$. Because the VRS PPS is the most and the CRS PPS is the least restrictive, the measured technical efficiency will be the highest under VRS and lowest under CRS. The frontiers of the CRS and NIRS production possibility sets coincide in the region of IRS. Similarly, the VRS and NIRS frontiers are identical in the DRS region.

Therefore, when IRS holds at (x^0, y^0) , in an input-oriented model $\theta_*^C = \theta_*^{NIRS} < \theta_*^V$ where the superscripts C, N, and V refer to CRS, NIRS, and VRS. Similarly, $\theta_*^C < \theta_*^{NIRS} = \theta_*^V$ implies DRS. Of course, in the case of CRS, all three estimates of technical efficiency equal unity. Note that in this nesting approach the possibility of multiple optimal solutions does not pose any problem because the objective function value does not differ across alternative optimal solutions.

6.4 Returns to Scale Properties of an Inefficient Input-Output Bundle

As has been noted earlier, returns to scale is a meaningful concept only when the input-output bundle is efficient and is a point on the frontier of the PPS. In the foregoing discussion it was assumed that (x^0, y^0) is technically efficient. When that is not the case, one must first project it on to the frontier and only then can examine the local RTS properties at the efficient projection. This, however, creates an ambiguity because there would be two alternative projections – the input oriented $(\theta^* x^0, y^0)$ and the output oriented $(x^0, \phi^* y^0)$. The nature of local RTS can be different at the two different points.

At this point it would helpful to define for any given input-output pair the 2-dimensional conditional PPS:

$$T(x^0, y^0) = \{(\alpha, \beta) : (\beta x^0, \alpha y^0) \in T; \alpha, \beta \geq 0\}. \quad (61)$$

In terms of the transformation function

$$T(x^0, y^0) = \{(\alpha, \beta) : F(\beta x^0, \alpha y^0) \leq 0; \alpha, \beta \geq 0\}. \quad (62)$$

Similarly, the conditional graph of the technology is

$$G(x^0, y^0) = \{(\alpha, \beta) : F(\beta x^0, \alpha y^0) = 0; \alpha, \beta \geq 0\}. \quad (63)$$

One can think of the graph in (63) as the *ray production function*

$$\alpha = g(\beta) \quad (64)$$

The MPSS for (x^0, y^0) corresponds to the highest ray average productivity along (64).

The following lemma from Ray (2010) shows that when the production possibility set is convex, IRS holds at all scales smaller than the smallest MPSS. Similarly, DRS holds at all scales larger than the largest MPSS.¹⁰

Lemma: *For any convex productivity possibility set T , if there exist non-negative scalars α and β such that $\alpha > \beta > 1$, and both (\tilde{x}, \tilde{y}) and $(\beta\tilde{x}, \alpha\tilde{y}) \in G$, then $\gamma > \delta$ for every γ and δ such that $1 < \delta < \beta$ and $(\delta\tilde{x}, \gamma\tilde{y}) \in G$.*

Proof: Because (\tilde{x}, \tilde{y}) and $(\beta\tilde{x}, \alpha\tilde{y})$ are both feasible, by convexity of T , for every

$\lambda \in (0, 1)$, $((\lambda + (1 - \lambda)\beta)\tilde{x}, (\lambda + (1 - \lambda)\alpha)\tilde{y})$ is also feasible. Now select λ such that $\lambda + (1 - \lambda)\beta = \delta$.

Further, define $\mu = \lambda + (1 - \lambda)\alpha$. Using these notations, $(\delta\tilde{x}, \gamma\tilde{y}) \in T$. But, because $(\delta\tilde{x}, \gamma\tilde{y}) \in G$, $\gamma \geq \mu$. However, because $\alpha > \beta$, $\mu > \delta$. Hence, $\gamma > \delta$.

An implication of this lemma is that, when the production possibility set is convex, if the technology exhibits locally diminishing returns to scale at smaller input scale, it cannot exhibit increasing returns at a bigger input scale. This is easily understood in the single-input single-output case. When both x and y are scalars, average productivity at (\tilde{x}, \tilde{y}) is $\frac{\tilde{y}}{\tilde{x}}$ and at $(\beta\tilde{x}, \beta\tilde{y})$ is $\frac{\alpha}{\beta} \frac{\tilde{y}}{\tilde{x}}$. Thus, when $\alpha > \beta$, average productivity has increased. The above lemma implies that for every input level x in between \tilde{x} and $\beta\tilde{x}$, average productivity is greater than $\frac{\tilde{y}}{\tilde{x}}$. Thus, average productivity could not first decline and then increase as the input level increased from \tilde{x} and $\beta\tilde{x}$.

Two results follow immediately. First, locally increasing returns to scale holds at every input-output bundle $(x, y) \in G$ that is smaller than the smallest MPSS. Second, locally diminishing returns to scale holds at every input-output bundle $(x, y) \in G$ that is greater than the largest MPSS. To see this, let $x = bx^*$ and $y = ay^*$, where (x^*, y^*) is the smallest MPSS for the given input and output mix. Because (x, y) is not an MPSS, $\frac{a}{b} < 1$. Further, assume that $b < 1$. Define $\beta = \frac{1}{b} (> 1)$ and $\alpha = \frac{1}{a}$. Then $(x^*, y^*) = (\beta x, \alpha y)$ and $\frac{\alpha}{\beta} > 1$. Because ray average productivity is higher at a larger input scale, by virtue of the lemma, locally increasing returns to scale holds at (x, y) . Next assume that $b > 1$. Again, because (x, y) is not an MPSS, $\frac{a}{b} < 1$. That is ray average productivity has fallen as the input scale is increased from x^* to $x = bx^*$. Then,

¹⁰ For a different proof see Banker and Thrall (1992).

by virtue of the lemma, ray average product could not be any higher than $\frac{a}{b}$ at a slightly greater input scale, $\bar{x} = (1 + \varepsilon)x$. But, because (x, y) is not an MPSS, ray average product cannot remain constant as the input scale is slightly increased. Hence, ray average product must fall as the input scale is slightly increased from x . Thus, locally diminishing returns to scale holds at every $(x, y) \in G$, when x is larger than the largest MPSS.

6.5 Finding the MPSS

One can solve the following optimization problem proposed by Cooper, Thompson, and Thrall (1996) to find the MPSS for the input-output bundle (x^0, y^0) :

$$\begin{aligned}
 & \max \frac{\alpha}{\beta} \\
 & s.t. \quad \sum_{j=1}^N \lambda_j y^j \geq \alpha y^0; \\
 & \quad \sum_{j=1}^N \lambda_j x^j \leq \beta x^0; \\
 & \quad \sum_{j=1}^N \lambda_j = 1; \\
 & \quad \alpha, \beta, \lambda_j \geq 0, (j = 1, 2, \dots, N).
 \end{aligned} \tag{65}$$

Even though the objective function is nonlinear, it can be easily transformed into a linear programming problem by defining $t = \frac{1}{\beta}$ and $\mu_j = t\lambda_j$ ($j = 1, 2, \dots, N$). Note that non-negativity of β and λ_j s ensures that t and μ_j s are also non-negative. Problem (65) can, therefore, be reformulated as the following linear programming problem:

$$\begin{aligned}
 & \max \rho \\
 & s.t. \quad \sum_{j=1}^N \mu_j y^j \geq \rho y^0; \\
 & \quad \sum_{j=1}^N \mu_j x^j \leq \beta x^0; \\
 & \quad \sum_{j=1}^N \mu_j = t; \\
 & \quad t, \mu_j \geq 0, (j = 1, 2, \dots, N).
 \end{aligned} \tag{66}$$

From the optimal solution of this problem we can derive $\beta^* = \frac{1}{t^*}$ and $\alpha^* = \frac{\rho^*}{t^*}$. One can then infer the nature of returns to scale from these values of α^* and β^* . It may be pointed out here that because the only restriction

on t is non-negativity, (67) is simply the output-oriented CCR DEA problem and $\frac{1}{\rho^*}$ is the same as the output-oriented CRS technical efficiency $\tau_y^C(x^0, y^0)$. Thus, (67) is in reality a minor modification of Banker (1984).

Because (x^0, y^0) is a feasible input-output bundle, $(\alpha = \beta = \rho = 1)$ is a feasible solution for this problem. Hence, the optimal value ρ^* is always greater than or equal to 1. When $\rho^* = \frac{\alpha^*}{\beta^*}$ exceeds unity, we know that (x^0, y^0) is *not an MPSS*. But, we can also conclude that $(\beta^* x^0, \alpha^* y^0)$ is *an MPSS*. When the bundle (x^0, y^0) is not itself an MPSS, $\rho^* > 1$ so that $\alpha^* > \beta^*$.

We may invoke the lemma above to determine the local RTS properties at (x^0, y^0) by comparing it with its MPSS. Note that in the conditional PPS (61), (x^0, y^0) corresponds to $(\alpha, \beta) = (1, 1)$. Let the MPSS be designated as $(\beta^* x^0, \alpha^* y^0) = (x^*, y^*)$. If the MPSS is unique, there are five different possibilities:

- (i) $1 < \beta^* < \alpha^*$; (ii) $\beta^* < \alpha^* < 1$;
- (iii) $\beta^* < 1 < \alpha^*$; (iv) $\beta^* = 1 < \alpha^*$, (v) $\beta^* < 1 = \alpha^*$.

In case (i), $x^0 < \beta^* x^0 = x^*$ and $y^0 < \alpha^* y^0 = y^*$. Thus, (x^0, y^0) lies towards the southwest of the MPSS.

Both input- and output-oriented projections of the bundle (x^0, y^0) fall in the region of IRS. In this case, the unit is conclusively too small relative to its MPSS. Similarly, in case (ii)

$x^0 > \beta^* x^0 = x^*$ and $y^0 > \alpha^* y^0 = y^*$ and both input- and output-oriented projections fall in the region of DRS. The implication is that the unit is too large. In case (iii), the RTS characterization depends on the direction of projection. Because $\beta^* < 1$, $x^* < x^0$, and the input scale is bigger than the MPSS. The output-oriented projection falls in the region of DRS. At the same time, because $1 < \alpha^*$, $y^o < y^*$ the output-scale is smaller than the MPSS and the input-oriented projection falls in the region of IRS. In case (iv)

$x^0 = x^*$ but $y^0 < y^*$. This time the input scale corresponds to the MPSS but the output scale is too small. Eliminating output oriented technical inefficiency will automatically project the observed bundle on to its MPSS. Similarly in case (v) $y^0 = y^*$ but $x^0 > x^*$. The input oriented efficient projection is the MPSS.

Figure 4 shows how the local RTS properties of an input-output bundle can be determined by comparing it to its MPSS. In this diagram the scale of the input bundle x^0 is measured as β along the horizontal axis and the scale of the output bundle y^0 is measured as α up the vertical axis. The piecewise

connected line $ABCDE$ shows the pairs (α, β) for which the corresponding input-output bundle $(\beta x^0, \alpha y^0)$ is on the frontier of the PPS. Two points of special interest are $(\alpha, \beta) = (1, 1)$ and the point C , $(\alpha, \beta) = (\alpha^*, \beta^*)$. The former is the location of the observed bundle (x^0, y^0) and the latter is its MPSS. The local RTS properties of (x^0, y^0) depends on where $(\alpha, \beta) = (1, 1)$ is located relative to C in this diagram. This is true even when $(1, 1)$ lies on the $ABCDE$ line and is technically efficient. If $(1, 1)$ lies in area (1) to the southwest of C , both input and output oriented projections will be smaller than the MPSS and IRS holds unequivocally at (x^0, y^0) . If it lies in area (2) towards the northeast of C both projections will be larger than the MPSS and DRS holds at (x^0, y^0) . By contrast, area (3) is an inclusive region. Here the output oriented projection is larger than the MPSS implying DRS but the input oriented projection is smaller than the MPSS implying IRS. The unit is too small judged by its output scale but is too large when judged by the input scale. Case (iv) corresponds to points on the vertical line $C\beta^*$ while points on the horizontal line through C below the frontier correspond to case (v).

The Case of Multiple MPSS

Next we consider the possibility of multiple MPSS. This is depicted graphically in Figure 5. Here both C_1 and C_2 are MPSS and so are their convex combinations lying on the line segment connecting them. At C_1 , (α_1^*, β_1^*) is the smallest MPSS. Similarly, (α_2^*, β_2^*) at C_2 is the largest MPSS. It is obvious that when the problem in (66) has a unique optimal solution (in particular, t^* is unique), there cannot be multiple MPSS. For multiple optimal solutions, the largest $t^* = \sum_j \mu_j^*$ across all optimal solutions of (66) corresponds to the smallest MPSS, β_1^* . Similarly, β_2^* corresponds to the smallest $t^* = \sum_j \mu_j^*$ at an optimal solution.

Note that across all optimal solutions the value of the objective function is the same (ρ^*) . Hence, $\beta_1^* = \frac{1}{t_1^*}$, where

$$t_1^* = \max \sum_j \mu_j$$

s.t.

$$\sum_j \mu_j x^j \leq x^0;$$

$$\sum_j \mu_j y^j \geq \rho^* y^0; \quad (67)$$

$$\mu_j (j = 1, 2, \dots, N) \geq 0.$$

Similarly, $\beta_2^* = \frac{1}{t_2^*}$, where

$$t_2^* = \min \sum_j \mu_j$$

s.t.

$$\sum_j \mu_j x^j \leq x^0;$$

$$\sum_j \mu_j y^j \geq \rho^* y^0; \quad (68)$$

$$\mu_j (j = 1, 2, \dots, N) \geq 0.$$

Once β_1^* and β_2^* have been determined from (67) and (68), the corresponding values of α are readily obtained as $\alpha_1^* = \rho^* \beta_1^*$ and $\alpha_2^* = \rho^* \beta_2^*$.

As shown in Figure 5, the set of output-input scales (α, β) for which the input-output bundles $(\beta x^0, \alpha y^0)$ are feasible can be partitioned into six different regions defined below:

- (i) In region (1) towards the southwest of the smallest MPSS (C_1), $(\beta < \beta_1^*; \alpha < \alpha_1^*)$. When (x^0, y^0) falls in this region, $1 < \beta_1^* < \alpha_1^*$. Hence, increasing returns to scale holds unambiguously.
- (ii) In region (2) to the northeast of the largest MPSS (C_2), $(\beta_2^* < \beta; \alpha_2^* < \alpha)$. If (x^0, y^0) falls in this region, $\beta_1^* < \alpha_1^* < 1$. Diminishing returns to scale holds unambiguously in this region.
- (iii) In region (3), $\beta_1^* < \beta < \beta_2^*$ while $\alpha_1^* < \alpha < \alpha_2^*$. Points in this region lie between the smallest and the largest MPSS. It is interesting to note, that even if the point $(\alpha = 1, \beta = 1)$ is not technically efficient and lies below the $C_1 C_2$ line, both the input- and the output-oriented projection of the inefficient bundle will fall in the region of constant returns to scale. Thus, there is no scale inefficiency in this region even though there may be technical inefficiency.

- (iv) In region (4), $\beta_2^* < \beta; \alpha < \alpha_1^*$. When the actual input-output bundle lies here, $\beta_2^* < 1 < \alpha_1^*$. The input bundle x^0 is larger than the largest MPSS hence the output oriented projection falls in the area of diminishing returns. At the same time, the actual output bundle is smaller than the smallest MPSS. Hence, increasing returns to scale holds at the input oriented projection. Thus, returns to scale cannot be unambiguously defined at the actual input-output bundle.
- (v) In region (5a), $\beta_1^* < \beta < \beta_2^*$ but $\alpha < \alpha_1^*$. When the actual input-output bundle lies here, y^0 is smaller than the smallest MPSS and the input oriented projection falls in the area of increasing returns. At the same time, the actual input bundle lies between the smallest and the largest MPSS. Hence, constant returns to scale holds at the output oriented projection. Here also the returns to scale characterization depends on the orientation.
- (vi) In region (5b), $\beta_2^* < \beta$ while $\alpha_1^* < \alpha < \alpha_2^*$. When the actual input-output bundle lies here, x^0 is larger than the largest MPSS. Hence the output oriented projection falls in the area of diminishing returns. At the same time, the actual output bundle lies between the smallest and the largest MPSS. Hence, constant returns to scale holds at the input oriented projection. Here the input bundle is too large. But the actual output bundle, if produced from the technically efficient input bundle would correspond to an MPSS.

7. Non-radial Measures of Efficiency

A problem with the radial models both input and output oriented is that when slacks are present in the input or output constraints at the optimal solution of the relevant DEA problem, the extent of inefficiency is underestimated. In an output oriented model, for example, a positive slack in any output constraint represents the amount by which the concerned output can be further expanded beyond the common expansion factor (φ^*). In an extreme example, if there is no room to increase one specific output without increasing inputs even if the other outputs can be doubled, the radial efficiency measure of the firm under evaluation will be 100%.¹¹

One way out of this paradoxical situation is to consider a non-radial measure of technical efficiency that reflects the full potential for increasing in every output even though not equi-proportionately. Färe and

¹¹ In fact, CCR (1979) included a small penalty (ε) for slacks in their DEA models. However, because ε is assumed to be an arbitrarily small (non-Archimedean) number, it cannot be incorporated in any practical application.

Lovell (1978) introduced what they called a Russell measure of technical efficiency that rules out the presence of output slacks in an output-oriented model and input slacks in an input oriented model.¹²

7.1 Output Oriented Russell Measure

The output-oriented Russell measure of technical efficiency is the inverse of the maximum average expansion factor across the individual outputs and is measured as

$$RM_y(x^0, y^0) = \frac{1}{\rho_y(x^0, y^0)} \quad (69)$$

where

$$\begin{aligned} \rho_y(x^0, y^0) = \max \frac{1}{m} \sum_{r=1}^m \varphi_r \\ s.t. \sum_{j=1}^N \lambda_j y_r^j \geq \varphi_r y_r^0; (r = 1, 2, \dots, m); \\ \sum_{j=1}^N \lambda_j x_i^j \leq x_i^0; (i = 1, 2, \dots, n) \\ \sum_{j=1}^N \lambda_j = 1; \\ \varphi_r \geq 1; (r = 1, 2, \dots, m); \\ \lambda_j \geq 0; (j = 1, 2, \dots, N). \end{aligned} \quad (70)$$

It can be seen that the radial output-oriented model can be recovered from this non-radial model by imposing the additional restriction $\varphi_r = \varphi$ for each input r ($r = 1, 2, \dots, m$). This, the BCC output-oriented measure under VRS (or the CCR measure under CRS) is a special case of the Russell measure and

$$\tau_y(x^0, y^0) \geq RM_y(x^0, y^0). \quad (71)$$

7.2 Non-radial Russell Input Efficiency

Analogous to (70) above is the input oriented Russell efficiency measure is the arithmetic mean of the input specific contraction factors (θ_i) across all inputs

¹² Input slacks may still be present in an output oriented model and output slacks in an input oriented model.

$$\begin{aligned}
RM_x(x^0, y^0) &= \rho_x(x^0, y^0) = \min \frac{1}{n} \sum_{i=1}^n \theta_i \\
s.t. \sum_{j=1}^N \lambda_j y_r^j &\geq y_r^0; (r = 1, 2, \dots, m); \\
\sum_{j=1}^N \lambda_j x_i^j &\leq \theta_i x_i^0; (i = 1, 2, \dots, n) \\
\sum_{j=1}^N \lambda_j &= 1; \\
\theta_i &\leq 1; (i = 1, 2, \dots, n); \\
\lambda_j &\geq 0; (j = 1, 2, \dots, N).
\end{aligned} \tag{72}$$

Again, the radial input oriented technical efficiency is a restricted version of the Russell input-oriented measure with $\theta_i = \theta$ for all i and

$$\tau_x(x^0, y^0) \geq RM_x(x^0, y^0). \tag{73}$$

8. Graph Efficiency Measures

The alternative measures of efficiency considered so far are either output oriented or input oriented. In an output oriented model, an inefficient unit is projected on to the frontier by the maximum proportional expansion of all of its outputs, but reducing inputs is not an objective. Similarly, in an input oriented model, the objective is only to scale down all inputs proportionately as much as possible. In measuring graph efficiency, one seeks to achieve some reduction in inputs side by side with expansion of outputs. The problem in this case would be that unlike in an output oriented or an input oriented model, the direction of projection on to the frontier is arbitrary. Two popular measures of graph efficiency are the ones based on the Graph Hyperbolic Distance Function due to Färe, Grosskopf, and Lovell (FGL) (1985) and the Directional Distance Function introduced by Chambers, Chung, and Färe (CCF) (1996).

8.1 Graph Hyperbolic Efficiency

For the Graph Hyperbolic efficiency measure one selects the point on the frontier that lies on a rectangular hyperbola through the observed input-output bundle. In a single output single input case both the actual point (x_0, y_0) and the efficient point (x^*, y^*) satisfy the equation $xy = k$. This is shown in

Figure 6 where the point A represents the observed input-output bundle (x_0, y_0) and the point B on the production frontier is the efficient projection (x^*, y^*) . The level of efficiency is $\frac{1}{\delta^*} = \frac{Ox^*}{Ox_0} = \frac{Oy_0}{Oy^*}$.

To obtain the efficient projection in the multiple output case, one needs to solve the problem:

$$\begin{aligned}
& \max \delta \\
& s.t. \sum_{j=1}^N \lambda_j y_r^j \geq \delta y_r^0; (r = 1, 2, \dots, m); \\
& \sum_{j=1}^N \lambda_j x_i^j \leq \frac{1}{\delta} x_i^0; (i = 1, 2, \dots, n) \\
& \sum_{j=1}^N \lambda_j = 1; \\
& \lambda_j \geq 0; (j = 1, 2, \dots, N); \delta \text{ unrestricted}
\end{aligned} \tag{74}$$

The Graph Hyperbolic measure of efficiency is

$$\tau_{GH}(x^0, y^0) = \frac{1}{\delta^*}. \tag{75}$$

Note that the input constraints in (74) are non-linear. However, if one assumes CRS, one can define

$\mu_j = \delta \lambda_j$ and $\varphi = \delta^2$ to rewrite the model as

$$\begin{aligned}
& \max \varphi \\
& s.t. \sum_{j=1}^N \mu_j y_r^j \geq \varphi y_r^0; (r = 1, 2, \dots, m); \\
& \sum_{j=1}^N \mu_j x_i^j \leq x_i^0; (i = 1, 2, \dots, n) \\
& \mu_j \geq 0; (j = 1, 2, \dots, N); \varphi \text{ unrestricted}
\end{aligned} \tag{76}$$

In this case,

$$\tau_{GH}^C(x^0, y^0) = \frac{1}{\sqrt{\varphi^*}}. \tag{77}$$

In the case of VRS, the problem in (72) remains non-linear. FGL linearize the input constraints using the Taylor's series approximation

$$\begin{aligned}
f(\delta) &= \frac{1}{\delta} \approx f(\delta_0) + f'(\delta_0)(\delta - \delta_0) \\
&= \frac{2-\delta}{\delta_0}.
\end{aligned} \tag{78}$$

Using $\delta_0 = 1$ as the point of approximation, the linearized version of (74) is

$$\begin{aligned}
&\max \delta \\
&s.t. \sum_{j=1}^N \lambda_j y_r^j \geq \delta y_r^0; (r = 1, 2, \dots, m); \\
&\quad \sum_{j=1}^N \lambda_j x_i^j + \delta x_i^0 \leq 2x_i^0; (i = 1, 2, \dots, n) \\
&\quad \sum_{j=1}^N \lambda_j = 1; \\
&\quad \lambda_j \geq 0; (j = 1, 2, \dots, N); \delta \text{ unrestricted}
\end{aligned} \tag{79}$$

Note that assuming $\delta_0 = 1$ amounts to assuming that the observed pint is on the frontier. When this is not the case, the approximation will be rather poor. Ideally, one should use $\delta_0 = 1$ only as the starting point and iterate (78)-(79) until convergence.

8.2 Directional Distance Function

Building upon Luenberger's (1992) *benefit function*, Chambers, Chung, and Färe (1996) introduced the Directional Distance Function (DDF) to measure the distance of an observed input-output bundle from the frontier of the PPS in a direction chosen by the analyst. Let $g^x = (g_1^x, g_2^x, \dots, g_n^x) \in R_+^n$ and

$g^y = (g_1^y, g_2^y, \dots, g_m^y) \in R_+^m$ be two direction sub-vectors. Then the DDF can be defined as

$$\vec{D}(x^0, y^0; g^x, g^y) = \max \beta : (x^0 - \beta g^x, y^0 + \beta g^y) \in T. \tag{80}$$

It is clear that one can recover the radial output oriented model by setting $g^x = 0$ and $g^y = y^0$. In that case, β in (80) would equal $(\varphi - 1)$ in (18) or (22) depending on the returns to scale assumption.

An interesting choice of the direction for projection would be $(g^x, g^y) = (x^0, y^0)$. That leads to,

$$\vec{D}(x^0, y^0; g^x, g^y) = \max \beta : ((1 - \beta)x^0, (1 + \beta)y^0) \in T. \tag{81}$$

In that case β is the maximum percentage by which all outputs can be expanded and all inputs can be contracted simultaneously.

In Figure 7, A is the observed bundle (x_0, y_0) . The point $B(g^x = -x^0, g^y = y_0)$ defines the direction of movement. The point C on the production frontier shows the maximum feasible movement within the production possibility set in the direction parallel to OB . In this case, the Directional Distance Function is $\beta = \frac{AC}{OB} = \frac{OD}{OB}$.

For an arbitrary choice of (g^x, g^y) , the relevant VRS DEA problem will be

$$\begin{aligned}
& \max \beta \\
& s.t. \sum_{j=1}^N \lambda_j y_r^j - \beta g_r^y \geq y_r^0; (r = 1, 2, \dots, m); \\
& \sum_{j=1}^N \lambda_j x_i^j + \beta g_i^x \leq x_i^0; (i = 1, 2, \dots, n) \\
& \sum_{j=1}^N \lambda_j = 1; \\
& \lambda_j \geq 0; (j = 1, 2, \dots, N); \beta \text{ unrestricted}
\end{aligned} \tag{82}$$

The flexibility of the DDF is apparent from the fact that it can be radial (setting $g^x = 0$ or $g^y = 0$), bi-radial (setting $g^x = x^0$ and $g^y = y^0$), or completely non radial for arbitrary choice of (g^x, g^y) .

Ray (2007) introduced a measure of overall inefficiency as

$$\varepsilon(x^0, y^0) = \max(\varphi - \theta) : (\theta x^0, \varphi y^0) \in T. \tag{83}$$

In a radial output oriented model, a measure of technical inefficiency is $\varphi - 1$ where φ is the maximum scaling factor for all outputs. Similarly, the input oriented technical inefficiency is $1 - \theta$ where θ is the minimum scaling factor for all inputs. In that sense, the overall inefficiency is the sum of output and input inefficiencies.

The DEA LP problem for (83) is

$$\begin{aligned}
& \max \varphi - \theta \\
& s.t. \sum_{j=1}^N \lambda_j y_r^j \geq \varphi y_r^0; (r = 1, 2, \dots, m); \\
& \sum_{j=1}^N \lambda_j x_i^j \leq \theta x_i^0; (i = 1, 2, \dots, n) \quad (84) \\
& \sum_{j=1}^N \lambda_j = 1; \\
& \lambda_j \geq 0; (j = 1, 2, \dots, N); \quad \beta \text{ unrestricted}
\end{aligned}$$

The dual LP for (81) is

$$\begin{aligned}
& \min \pi \\
& s.t. \pi \geq \sum_{r=1}^m u_r y_r^j - \sum_{v=1}^n v_i x_i^j; (j = 1, 2, \dots, N) \\
& \sum_{r=1}^m u_r y_r^0 = 1; \sum_{i=1}^n v_i x_i^0 = 1; \quad (85) \\
& u_r \geq 0; v_i \geq 0; (r = 1, 2, \dots, m; i = 1, 2, \dots, n); \\
& \pi \text{ unrestricted.}
\end{aligned}$$

Ray (2007) has shown that if the optimal π^* in (85) is positive, then there does not exist any pair of non-negative shadow price vector (u, v) corresponding to which the bundle (x^0, y^0) would be the profit maximizing input-output pair.

Further, (84) can be seen to be a less restricted form of (81). Define $\beta^y = \varphi - 1$ and $\beta^x = 1 - \theta$. Then, (84) becomes

$$\begin{aligned}
& \max \beta^y + \beta^x \\
& s.t. \sum_{j=1}^N \lambda_j y_r^j \geq (1 + \beta^y) y_r^0; (r = 1, 2, \dots, m); \\
& \sum_{j=1}^N \lambda_j x_i^j \leq (1 - \beta^x) x_i^0; (i = 1, 2, \dots, n) \quad (86) \\
& \sum_{j=1}^N \lambda_j = 1; \\
& \lambda_j \geq 0; (j = 1, 2, \dots, N); \quad \beta^x, \beta^y \text{ unrestricted}
\end{aligned}$$

By imposing the restriction $\beta^x = \beta^y$ one gets the problem in (81) except for scaling the objective function.¹³

8.3 Pareto-Koopmans Measures

A Russell efficient output bundle contains no output slack. Similarly, no input slack can be present in a Russell efficient input bundle. However, to be Pareto-Koopmans efficient an input-output bundle must be simultaneously Russell efficient in both output and input orientations. Thus, Pareto-Koopmans (PK) efficiency combines both input and output oriented Russell efficiency. There are different variants of this PK efficiency but the most popular of them is the product of the Russell output and input efficiencies.¹⁴ It is called Enhanced Russell Measure by Pastor, Louis, and Sirvent (PLS) (1999), Slack based Measure (SBM) by Tone (2001),¹⁵ and simply Pareto-Koopmans efficiency by Ray (2004) and can be measured as

$$\begin{aligned} \tau^{PK}(x^0, y^0) &= \min \frac{\frac{1}{n} \sum_i \theta_i}{\frac{1}{m} \sum_r \varphi_r} \\ \text{s.t. } \sum_j \lambda_j y_{rj} &\geq \varphi_r y_{r0}; (r = 1, 2, \dots, m) \\ \sum_j \lambda_j x_{ij} &\leq \theta_i x_{i0}; (i = 1, 2, \dots, n) \\ \varphi_r &\geq 1; (r = 1, 2, \dots, m); \\ \theta_i &\leq 1; (i = 1, 2, \dots, n) \\ \sum_j \lambda_j &= 1; \lambda_j \geq 0; (j = 1, 2, \dots, N) \end{aligned} \quad (87)$$

Note that in (87) every input and output constraint will be strictly binding. Therefore at the optimal projection $x_i^* = \sum_j \lambda_j^* x_{ij} = \theta_i^* x_{i0}$ ($i = 1, 2, \dots, n$). Define the total reduction in input i as for each input

$s_i^- = x_i^0 - x_i^* \geq 0$. This leads to

$$\theta_i^* = \frac{x_i^*}{x_{i0}} = 1 - \frac{s_i^-}{x_{i0}}. \quad (88)$$

Similarly by defining $s_r^+ = y_r^* - y_{r0}$, we can derive

¹³ The model in (86) is further developed in Aparicio, Pastor, and Ray (2013).

¹⁴ Portela and Thanassoulis (2005) used the measure $\frac{\Pi\left(\frac{1}{\theta_i^n}\right)}{\Pi\left(\frac{1}{\varphi_r^m}\right)}$ and called it the Geometric Distance Function.

¹⁵ Tone's SBM appeared in 2001 in EJOR but makes no reference to the PLS Enhanced Russell measure introduced in the same journal in 1999 and the two are virtually identical.

$$\varphi_r^* = \frac{x_i^*}{x_{io}^*} = 1 + \frac{s_r^+}{y_{r0}}, (r = 1, 2, \dots, m) \quad (89)$$

Hence the objective function in (87) becomes

$$\frac{\frac{1}{n} \sum_i \theta_i}{\frac{1}{m} \sum_r \varphi_r} = \frac{1 - \frac{1}{n} \sum_i \frac{s_i^-}{x_{i0}}}{1 + \frac{1}{m} \sum_r \frac{s_r^+}{y_{r0}}} \quad (90)$$

which is the SBM.¹⁶

Both PLS and Tone use the expression in (90) for the objective function and resort to a normalization to convert the linear fractional functional programming problem into an LP following Charnes and Cooper (1968). Ray (2004), Ray and Jeon (2009), and Ray and Ghose (2014) on the other hand, use a linear approximation of the objective function at $(\theta_i = 1, \varphi_r = 1) (i = 1, 2, \dots, n; r = 1, 2, \dots, m)$ to get

$$\frac{\frac{1}{n} \sum_i \theta_i}{\frac{1}{m} \sum_r \varphi_r} \approx 2 + \frac{1}{n} \sum_i \theta_i - \frac{1}{m} \sum_r \varphi_r \quad (91)$$

and used $\min \frac{1}{n} \sum_i \theta_i - \frac{1}{m} \sum_r \varphi_r$ as the objective function. At the optimal solution of (87), one can

multiplicatively decompose the overall pareto-Koopmans efficiency as

$$\tau_{PK}(x^0, y^0) = PK_x(x^0, y^0) \cdot PK_y(x^0, y^0) \quad (92)$$

where

$$PK_x(x^0, y^0) = \frac{1}{n} \sum_i \theta_i^* \quad (93)$$

is the input efficiency and

$$PK_y(x^0, y^0) = \frac{1}{\frac{1}{m} \sum_r \varphi_r^*} \quad (94)$$

is the output efficiency of the firm.

9. Conclusion

¹⁶ A somewhat different measure of Pareto-Koopmans efficiency is the Range Adjusted measure (RAM) introduced by Cooper, Park, and Pastor (1999).

Over the decades since it was introduced in the Operations Research literature, new DEA models have been formulated to measure cost and profit efficiency as well as to test various characteristics of the technology like productivity change over time, capacity utilization, benefits of merger or break up of firms among many other areas of application in economics. In fact, DEA can now legitimately be described as a full blown nonparametric approach to production analysis and goes way beyond merely evaluating technical efficiency.¹⁷ Given its scope, this chapter deals only with measurement of different kinds of technical efficiency. Finally, as noted at the beginning one can generate empirical distributions of the frontier output at each input bundle through bootstrapping to create upper and lower bounds on the measured technical efficiency of each firm in the sample.¹⁸

¹⁷ For more detailed exposition of DEA as a nonparametric approach to neoclassical production economics see Färe, Grosskopf, and Lovell (1994) and Ray (2004). Cooper, Seiford, and Tone (2002) deals with most of the topics from an OR perspective.

¹⁸ For details on how to bootstrap for DEA see Simar and Wilson (1998).

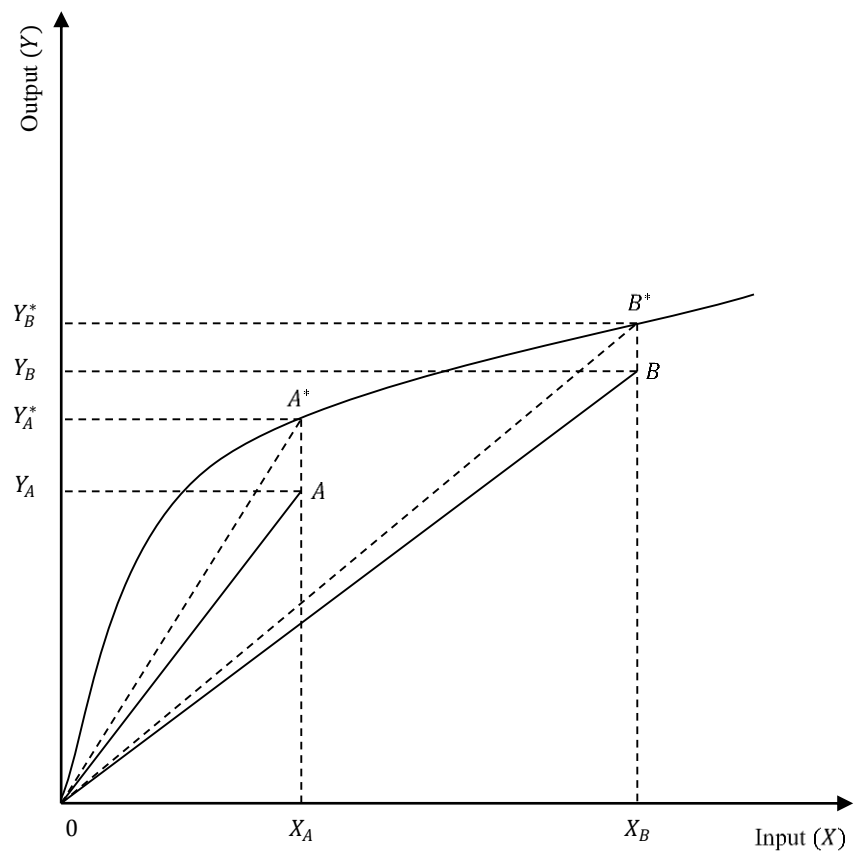


Figure 1: Output-oriented Technical Efficiency

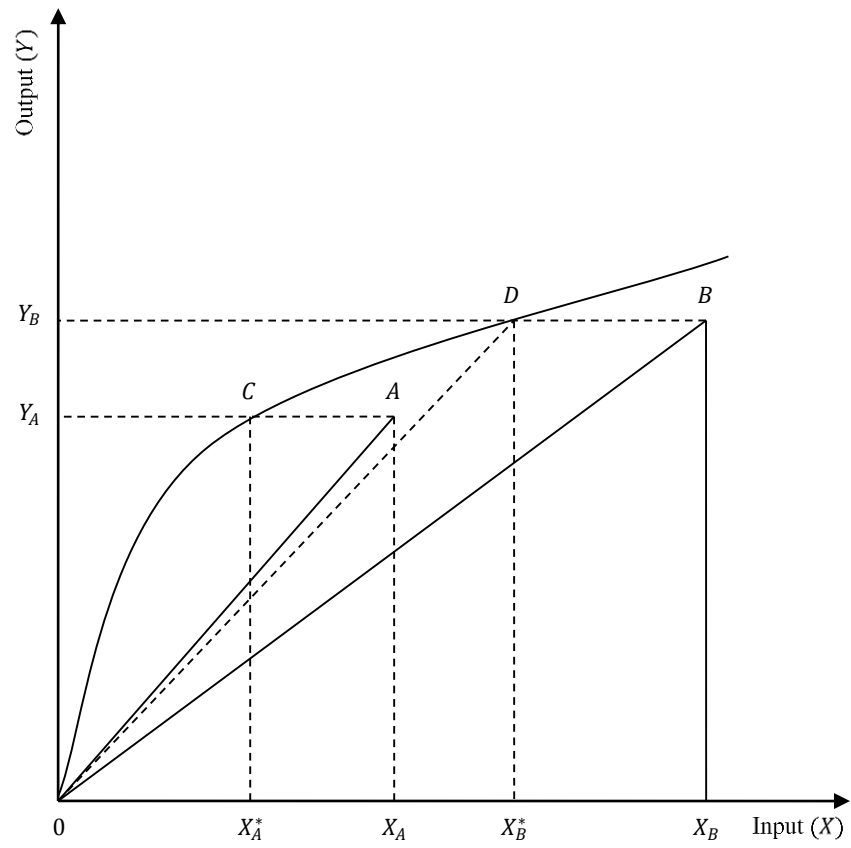


Figure 2: Input-oriented Technical Efficiency

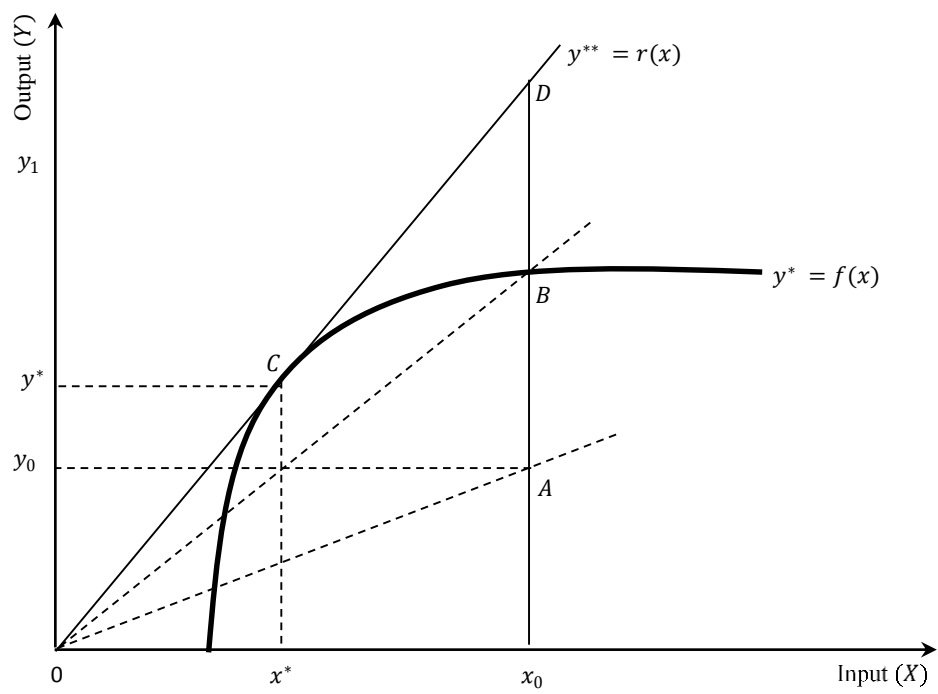


Figure 3: Scale Efficiency

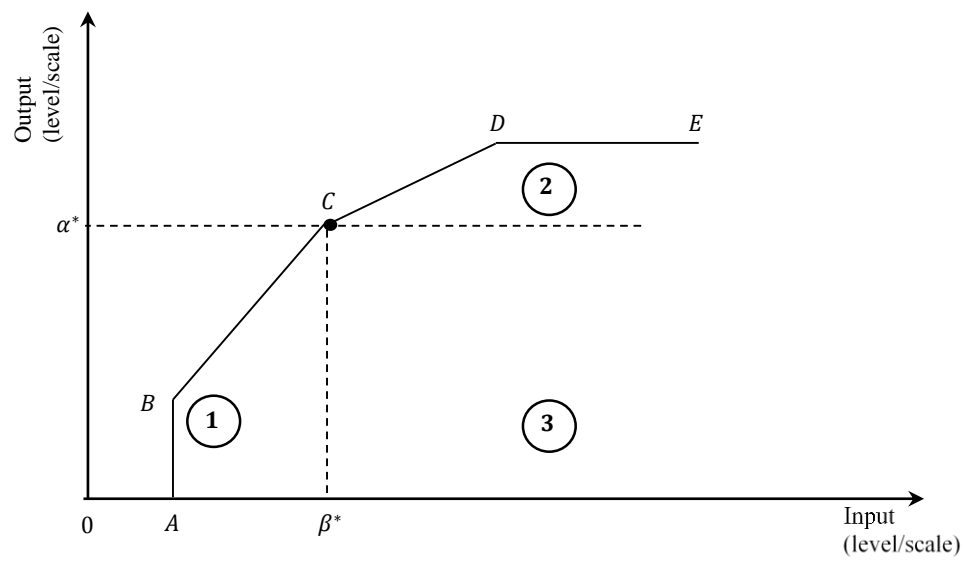


Figure 4: MPSS and RTS Regions

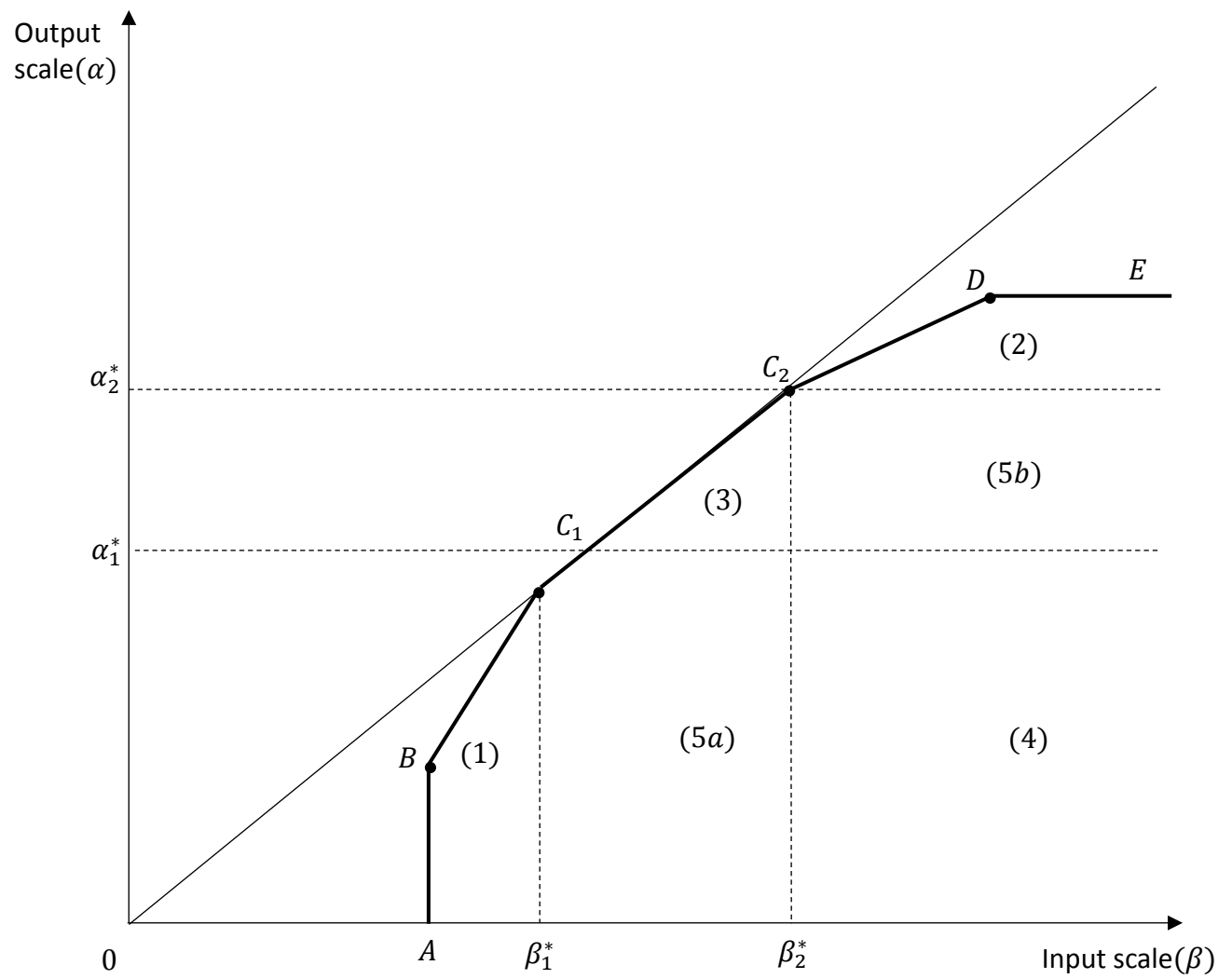


Figure 5
Multiple MPSS and the regions of the increasing, decreasing and the ambiguous returns to scale

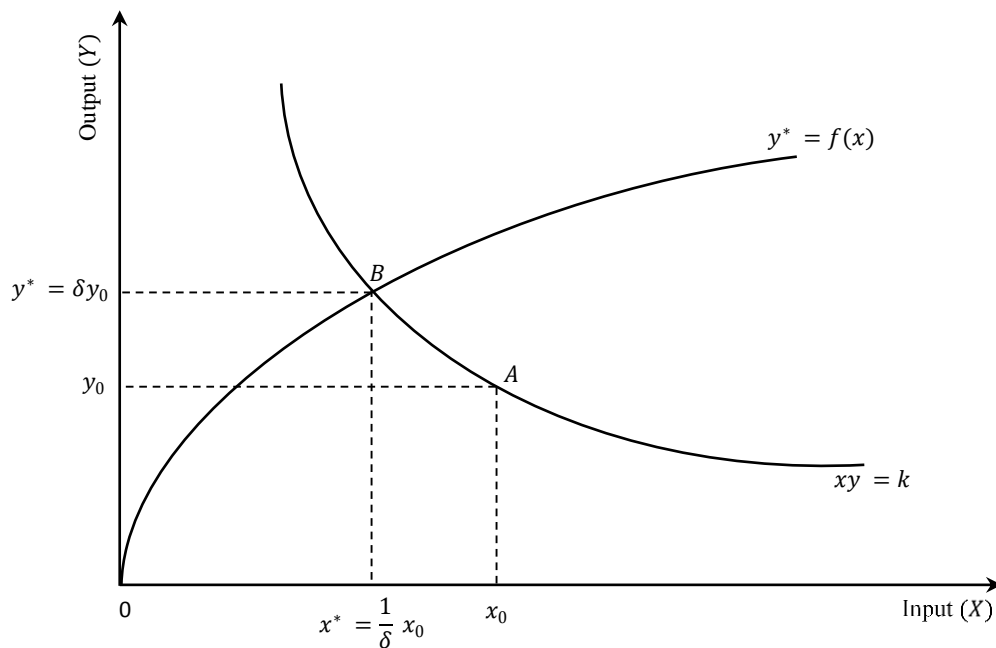


Figure 6: Graph Hyperbolic Distance Function

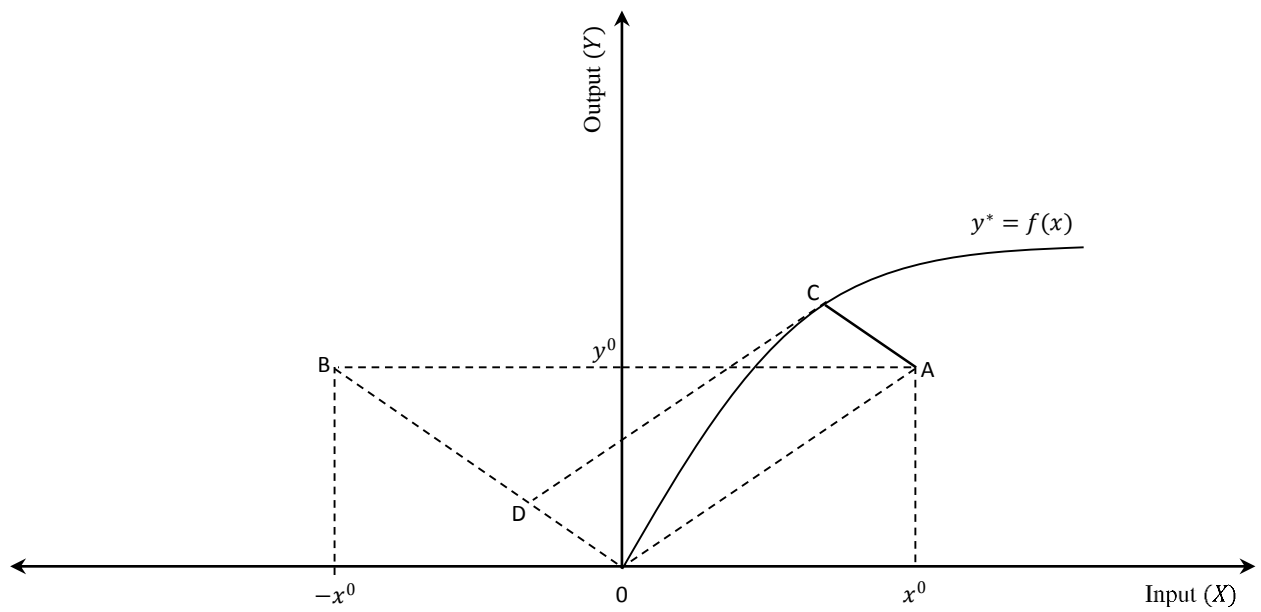


Figure 7: Directional Distance Function

References:

- Afriat, S. (1972) "Efficiency Estimation of Production Functions"; *International Economic Review*; 13:3; 568-98.
- Aigner, D.J. and S. F.Chu (1968), "On Estimating the Industry Production Function," *American Economic Review* 58:4 (September) 826-39.
- Aigner, D.J., C.A.K. Lovell, and P. Schmidt (1977) "Formulation and Estimation of Stochastic Frontier Production Function Models"; *Journal of Econometrics*, 6:1, 21-37.
- Aparicio, J., J.T. Pastor, and S.C. Ray (2013) "An overall measure of technical inefficiency at the firm and at the industry level: The 'lost profit on outlay'"; *European Journal of Operational Research*, Volume 226, Issue 1, 154-162
- Banker, R.D. (1984), "Estimating the Most Productive Scale Size Using Data Envelopment Analysis", *European Journal of Operational Research* 17: 1 (July) 35-44.
- Banker, R.D. and A. Maindiratta (1988) "Nonparametric Analysis of Technical and Allocative Efficiencies in Production", *Econometrica*, (November), pp. 1315-1332.
- Banker RD. (1993) "Maximum Likelihood, Consistency, and Data Envelopment Analysis: A Statistical Foundation" *Management Science*; 39; 1265-1273.
- Banker, R.D., A. Charnes, and W.W. Cooper (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Science*, 30:9 (September), 1078-92.
- Banker, R.D., and R. Natarajan (2008) "Evaluating Contextual Variables Affecting Productivity using Data Envelopment Analysis". *Operations Research*, Volume 56, Issue 1, pp. 48-58.
- Banker, R.D. and R.M. Thrall (1992) "Estimating Most Productive Scale Size using Data Envelopment Analysis", *European Journal of Operational Research*, 62, 74-84.
- Chambers, R.G., Y. Chung, and R. Färe (1996) "Benefit and Distance Functions", *Journal of Economic Theory*, 70 (August 1996), 407-19.

- Chambers, R.G., Chung, Y., Färe, R. (1998) “Profit, directional distance functions, and nerlovian Efficiency”. *Journal of Optimization Theory and Applications* 98, 351– 364
- Charnes, A. and W.W. Cooper (1968) “Programming with Linear Fractional Functionals”, *Naval Research Logistics Quarterly* 15; 517-522.
- Charnes, A., W.W. Cooper, and E. Rhodes (1978) “Measuring the Efficiency of Decision Making Units,” *European Journal of Operational Research* 2:6 (November), 429-44.
- Charnes, A., W.W. Cooper, and E. Rhodes (1979) “Short Communication: Measuring the Efficiency of Decision Making Units,” *European Journal of Operational Research* 3:4, 339
- Cooper, W.W., S.K. Park, K.S., and J.T. Pastor (1999) RAM: A Range Adjusted Measure of Inefficiency for Use with Additive Models, and Relations to Other Models and Measures in DEA; *Journal of Productivity Analysis*, 11, 5–42
- Cooper, W.W., L. Seiford and K. Tone.(2000) *Data Envelopment Analysis: A Comprehensive Text with Uses, Example Applications, References and DEA-Solver Software*. (Norwell, Mass: Kluwer Academic Publishers.)
- Cooper, W.W., R.G. Thompson, and R.M. Thrall (1996) “Introduction: Extensions and New Developments in DEA”; *Annals of Operations Research*, 66, pp 3-45.
- Diewert, E. and C. Parkan (1983) “Linear Programming Tests of Regularity Conditions for Production Functions” in W. Eichorn, R. Henn, K. Neumann, and R.w. Shephard, eds, *Quantitative Studies in production and Prices* (Würzburg: Physica-Verlag)
- Färe, R. and C.A.K. Lovell (1978), “Measuring the Technical Efficiency of Production,” *Journal of Economic Theory* 19:1 (October), 150-62.
- Färe, R., S. Grosskopf, and C.A.K. Lovell (1985) *The Measurement of Efficiency of Production* Boston: Kluwer-Nijhoff.

- Färe, R., S. Grosskopf, and C.A.K. Lovell (1994) *Production Frontiers* Cambridge: Cambridge University Press.
- Färe, R., S. Grosskopf, M. Norris, and Z. Zhang. “Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries”, *American Economic Review*, 84 (1994),66-83.
- Färe R, Grosskopf S, Lovell CAK, Yaisawarng S (1993) Derivation of shadow prices for undesirable outputs: a distance function approach. *The Review of Economics and Statistics* 75: 374-380.
- Färe, R., S. Grosskopf, and C.A.K. Lovell, and C. Pasurka (1989) “Multilateral Productivity Comparisons when Some Outputs are Undesirable: A Non-parametric Approach”, *Review of Economics and Statistics*,71:1 (February) 90-98.
- Farrell, M.J. (1957), “The Measurement of Technical Efficiency,”, *Journal of the Royal Statistical Society* Series A, General, 120, Part 3, 253-81.
- Farrell, M.J. and M. Fieldhouse (1962), “Estimating Efficient Production Functions Under Increasing Returns to Scale”, *Journal of the Royal Statistical Society* Series A, General, 125, Part 2, 252-67.
- Frisch, R. (1965) *Theory of Production*. Chicago: Rand McNally and Company.
- Greene, W. (1980), “Maximum Likelihood Estimation of Econometric Frontier Functions,” *Journal of Econometrics*, 13, pp. 27–56.
- Hanoch, G. and M. Rothschild (1972) “Testing the Assumption of Production Theory: A Nonparametric Approach”; *Journal of Political Economy*; 80:2 (March/April) 256-75.
- Kerstens, K. and P. Vanden Eeckaut (1999) “Estimating Returns to Scale Using Nonparametric Deterministic Technologies: A New Method Based on Goodness-of-Fit”; *European Journal of Operational Research*, 113(1), 206-214.
- Kumbhakar, S. and C.A.K. Lovell (2000) *Stochastic Frontier Analysis* (New York: Cambridge University Press).

- Luenberger, D.G. (1992) “Benefit Functions and Duality”, *Journal of Mathematical Economics*, 21, 115-145.
- Pastor, J.T., J.L. Ruiz, and I. Sirvent (1999) “An Enhanced DEA Russell-Graph Efficiency Measure”, *European Journal of Operational Research*, 115, 596-607.
- Portela, M.C.A.S., and Thanassoulis, E. (2005) “Profitability of a sample of Portuguese bank branches and its decomposition into technical and allocative components”, *European Journal of Operational Research* 162/3, 850–866.
- Ray, S.C. (2004) *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research* (New York: Cambridge University Press).
- Ray, S.C. (2007) Shadow profit maximization and a measure of overall inefficiency. *Journal of Productivity Analysis* 27, 231–236.
- Ray, S.C. (2010) “A One-Step Procedure for Returns to Scale Classification of Decision Making Units in Data Envelopment Analysis” *University of Connecticut Economics Working Paper* 2010-07.
- Ray, S.C. (2016) “Data Envelopment Analysis: An Overview”; in R.N. Sengupta, A. Gupta, and J. Datta (Eds) *Decision Sciences: Theory and Practice* (Taylor and Francis) pp 957-997.
- Ray, S.C. (2017) “The Transformation Function, Technical Efficiency, and the CCR Ratio”; *University of Connecticut Economics Working Paper* 2017-09.
- Ray, S.C. and A. Ghose (2014) “Production efficiency in Indian agriculture: An assessment of the post green revolution years”; *Omega* 44(2014)58–69
- Ray, S.C. and Y. Jeon (2009) “Reputation and efficiency: A non-parametric assessment of America’s top-rated MBA programs”; *European Journal of Operational Research* 189 (2008) 245–268
- Richmond, J., 1974, “Estimating the Efficiency of Production,” *International Economic Review*, 15, pp. 515–521.
- Schmidt, P., 1976, “On the Statistical Estimation of Parametric Frontier Production Functions,” *Review of*

Economics and Statistics, 58, pp. 238–239

- Seiford, L. and J. Zhu. “An investigation of returns to scale in data envelopment analysis”; *Omega, Int. J. Management Science* 27 (1999) 1-11
- Shephard, R. W. (1953), *Cost and Production Functions* Princeton: Princeton University Press.
- Shephard, R.W. (1970), *Theory of Cost and Production Functions* (Princeton)
- Simar, L. and P. Wilson (1998) “Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models”, *Management Science*, 44:11, 49-61
- Starrett, D.A. (1977) “Measuring Returns to Scale in the Aggregate, and the Scale Effect of Public Goods”; *Econometrica* , 45:6, 1439-1455.
- Tone, K. (2001) A slacks-based measure of efficiency in data envelopment analysis, *European Journal of Operational Research* 130, 498–509.
- Varian, H. R.(1984), “The Nonparametric Approach to Production Analysis,” *Econometrica* 52:3 (May) 579-97.
- Zhu, J. (2003) *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver* (Boston: Kluwer Academic Press).