University of Wisconsin-Madison

February, 1999 **Staff Paper No. 422**

A GENERALIZED DISTANCE FUNCTION AND THE ANALYSIS OF PRODUCTION EFFICIENCY

By

Jean Paul Chavas
and
Thomas L. Cox

AGRICULTURAL & APPLIED ECONOMICS

STAFF PAPER SERIES

A GENERALIZED DISTANCE FUNCTION

AND THE ANALYSIS OF PRODUCTION EFFICIENCY

by

Jean-Paul Chavas

and

Thomas L. Cox[1]

Abstract: A generalization of Shephard's distance functions is proposed, extending the usefulness of distance functions in economic analysis. Applications to efficiency measurements and productivity analysis are presented. New indexes of productivity, and technical, allocative and scale efficiency are proposed and analyzed. Interpretation of these indexes in terms of ray-average cost, ray-average revenue, and cost-to-revenue ratio is discussed.

Keywords: distance functions, efficiency, productivity.

J.E.L. classification number: D2, O3.

---

[1] Professors of Agricultural and Applied Economics, University of Wisconsin, Madison. We would like to thank Rolf Färe and two anonymous reviewers for useful comments on an earlier draft of this paper. This research was funded in part by a Hatch grant.

A GENERALIZED DISTANCE FUNCTION AND THE ANALYSIS OF PRODUCTION EFFICIENCY

1- INTRODUCTION

The measurement of productivity and efficiency has been a topic of considerable interest in economics. Much research has focused on the analysis of technical, allocative and scale efficiency of production activities (e.g., Debreu; Farrell; Farrell and Fieldhouse; Afriat; Forsund et al.; Russell; Färe et al., 1985, 1994a; Banker and Maindiratta; Bauer; Seiford and Thrall; Forsund et al.; Bauer; Seiford and Thrall), and the analysis of technological progress and productivity (e.g., Diewert; Caves et al., 1982a, 1982b). Farrell's efficiency indexes have been commonly used in empirical research on production efficiency for two reasons: they can be combined easily into an overall efficiency index; and they have an intuitive interpretation in terms of cost ratios or average cost ratios (Farrell; Farrell and Fieldhouse).

Shephard's distance functions have guided much of the development in productivity analysis and efficiency analysis. For example, Caves et al. (1982b) have investigated productivity indexes derived from Shephard's distance functions. And Färe et al. (1985, 1994a) have shown how the Farrell efficiency indexes are closely related to Shephard's distance functions. In a multi-input multi-output framework, Shephard defines two distance functions: an <u>input distance function</u> that rescales all inputs toward the frontier technology; and an <u>output distance function</u> that rescales all outputs toward the frontier. Unfortunately, unless the technology exhibits constant return to scale, these two distance functions differ and provide different measures of productivity and efficiency (Färe et al., 1985, 1994a; Caves et al., 1982b). This appears rather undesirable. Also, to be empirically meaningful, Shephard's distance functions rely on an "attainability assumption". This assumption states that all output vectors can be obtained from the rescaling of any non-zero input vector, or that all input vectors are feasible in the production of any rescaled non-zero output vector (see Shephard, chapter 9). However, in some situations, this attainability assumption may not be satisfied, especially if some inputs or outputs are not essential (see

Shephard; Färe and Mitchell). This can greatly limit the empirical usefulness of the methodology. To illustrate, consider Ray and Desli 's recent investigation of productivity growth and efficiency in industrialized countries. Using Shephard's output distance function, Ray and Desli were unable to report empirical estimates of technical change and scale efficiency for Ireland because the associated data did not satisfy the attainability assumption (Ray and Desli, p. 1037). This suggests a need to extend Shephard's distance functions.

Shephard's distance functions have been generalized in a number of ways. Graph measures of production efficiency have been developed by Färe et al. (1985, chapters 5-7; 1994a, chapter 8). For example, Färe et al. (1985, p. 110; 1994a, p. 198) defined a "Farrell graph technical efficiency index" that rescales both inputs and outputs equiproportionally. Other extensions of Farrell technical efficiency include a "generalized Farrell graph" measure proposed by Färe et al. (1985, p. 125), non-radial efficiency measures discussed by Russell and Färe et al. (1985, chapter 7), a "Farrell proportional distance" measure defined by Briec, and the shortage and benefit functions developed by Luenberger (1992, 1995). Briec's "Farrell proportional distance" function and Luenberger's shortage function are the same: they both allow the rescaling of inputs and outputs in <u>any</u> particular direction. As such, they provide a broad generalization to Shephard's distance functions (Chambers et al., 1996a, 1996b). They include as special cases most measures of technical efficiency found in the literature (Briec).

Thus it appears desirable to rescale inputs and outputs in a more flexible way than done in Shephard's distance functions. The Luenberger-Briec approach provides a general framework for doing so. However, it does not provide clear guidance for choosing the rescaling direction for inputs and outputs in efficiency analysis. Also, while the Farrell efficiency measures can be easily interpreted in terms of average cost, such interpretation is not straightforward in the Luenberger-Briec approach. This is somewhat unfortunate since average cost is a basic concept found in all production economic textbooks and commonly used in empirical economic analysis. This suggests considering a rescaling scheme for inputs and outputs that extends Shephard's distance functions while retaining the intuitive average cost interpretation of the Farrell indexes.

The objective of this paper is to propose a generalized Shephard's distance function with the following characteristics. First, it includes as special cases both Shephard's input and output distance functions while relaxing Shephard's attainability assumptions. Second, it generates efficiency indexes that have an intuitive interpretation in terms of average cost (or ray-average cost in a multi-output framework; see Baumol et al.). Third, these efficiency indexes can be combined easily into an overall efficiency index.

Our analysis is presented in a multi-input multi-output framework. Our generalized Shephard's distance function considers the simultaneous rescaling of both inputs and outputs. The direction of rescaling depends on a single parameter α that can vary between 0 and 1. As special cases, the parameter α taking the value 0 (1) implies only input (output) rescaling. Thus, our generalized distance function nests as special cases both Shephard's input and output distance functions. It applies without Shephard's attainability assumption, thus widening the range of applications of distance functions in economic analysis. Also, our proposed approach resolves the current dilemma concerning which Shephard's distance function (i.e., the input distance function, or the output distance function) to use when the technology departs from constant return to scale.[1]

The paper is organized as follows. Section 2 defines a generalized distance function and investigates its properties. Our generalized distance function provides a basis for investigating production efficiency and productivity growth. We propose new indexes of productivity, technical efficiency, allocative efficiency, and scale efficiency. Indexes of technical efficiency and productivity are presented in section 3. Our proposed technical efficiency index nests as special cases both the traditional input-based and output-based technical efficiency indexes commonly found in the literature (e.g., Färe et al., 1985, 1994a). Similarly, our productivity index nests as special cases both the input-based and output-based Malmquist productivity indexes discussed by Caves et al. (1982b). Our allocative efficiency indexes are presented in section 4. They are motivated from the cost function, from the revenue function, as well as from the profit function. Section 5 discusses our proposed scale efficiency indexes, based again on the cost function, the revenue function, as well as the profit function. In section 6, we show how all our proposed indexes can be conveniently interpreted in terms of the properties of ray-average cost, ray-average revenue, and cost-to-revenue ratios. This simple economic interpretation provides intuitive appeal to our proposed approach.

2- A GENERALIZED DISTANCE FUNCTION

Consider a production process involving a $(n\times1)$ input vector $x \in R_+^n$ used in the production of a $(m\times1)$ output vector $y \in R_+^m$.[2] The underlying technology is represented by the feasible set T, $(x, y) \in T \subseteq R_+^{m+n}$, or equivalently by the associated input requirement set $V(y, T) = \{x: (x, y) \in T, y \in R_+^m\}$. The following assumptions will be made throughout the paper:[3]

A1. The feasible set T is non-empty, and there exists an $x \geq 0$ and $y \geq 0$ such that $(x, y) \in T$.

A2. Nested: If $x \in V(y, T)$ and $y \geqq y'$, then $x \in V(y', T)$, $\forall\, y, y' \in R_+^m$.

A3. Monotonic: If $x \in V(y, T)$ and $x' \geqq x$, then $x' \in V(y, T)$, $\forall\, y \in R_+^m$.

A4. The feasible set T is closed.

Assumption A1 states that it is possible to produce some positive output from some positive input. Assumptions A2 and A3 have been called "strong disposability" of outputs and inputs, respectively. Assumption A4 implies the existence of isoquants at the boundary of the feasible set.

In general, we will assume that T represents a variable-return-to-scale (VRTS) technology. The nature of returns to scale can be characterized globally as follows.[4]

Definition 1: For all $y \geq 0$,

$$V(\lambda\, y, T) = \lambda\, V(y, T), \text{ for } \lambda > 0, \text{ under constant return to scale (CRTS);} \quad (1a)$$

$V(\lambda\, y, T) \subseteq (\supseteq)\ \lambda\, V(y, T)$, for $0 < \lambda \leqq 1$ $(\lambda \geqq 1)$, under increasing return to scale (IRTS); (1b)

$V(\lambda\, y, T) \supseteq (\subseteq)\ \lambda\, V(y, T)$, for $0 < \lambda \leqq 1$ ($\lambda \geqq 1$),

under decreasing return to scale (DRTS). (1c)

From (1a), under CRTS, a proportional change in all outputs is associated with the same proportional change in all inputs. In this case, the underlying production frontier is linearly homogeneous. From (1b) and (1c), under IRTS (DRTS), a proportional increase in all outputs is associated with a less than proportional (more than proportional) increase in all inputs. And a technology T characterized by variable return to scale (VRTS) can exhibit local IRTS and DRTS in different regions of the feasible space.

Shephard (Chapter 9) makes an attainability assumption: input attainability, which states that all output vectors can be obtained from the rescaling of any non-zero input vector; or output attainability, which says that all input vectors are feasible in the production of any rescaled non-zero output vector.[5] The implications of attainability for factor combinations have been investigated by Shephard (p. 45-50), and Färe and Mitchell. However, Shephard's attainability assumption is rather restrictive. For example, Färe and Mitchell have presented empirical evidence that the production technology of some U.S. manufacturing sectors does not satisfy Shephard's input attainability axiom.

Another example is given in the context of piece-wise linear representation of technology commonly used in data envelopment analysis (DEA) (e.g., Banker; Banker, Charnes and Cooper). Consider an input-output data set on a sample of K firms in a given industry: (x_k, y_k), k = 1, 2, ..., K. A piece-wise linear representation of the underlying technology is given by the specification:

$$T_o = \{(x, y): \Sigma_k(\lambda_k\, y_k) \geqq y,\ \Sigma_k(\lambda_\kappa\, x_k) \leqq x,\ \Sigma_k\, \lambda_k = 1,\ \lambda_k \in R^+,\ k = 1, 2, ..., K\}.$$

As shown by Afriat, Banker and Maindiratta, and others, the production possibility set T_o is an inner bound representation of the underlying technology under convexity and variable return to scale. In general, the set T_o commonly used in data envelopment analysis (e.g., Banker; Banker and Maindiratta; Banker, Charnes and Cooper) does not satisfy Shephard's input or output attainability axiom. This is illustrated in Figure 1 where it is clear that point B is not Shephard-attainable given that the production possibility set is bounded by (abcde).[6] This means that both Shephard's input distance function and his output distance function do not take finite values when evaluated at point B. This limits the empirical usefulness of Shephard's distance functions in the context of DEA analysis. This limitation is illustrated in Ray and Desli 's recent investigation of productivity growth and efficiency in industrialized countries. Using Shephard's output distance function, Ray and Desli were unable to report empirical estimates of technical change and scale efficiency for Ireland because the associated data did not satisfy the attainability assumption (Ray and Desli, p. 1037).

These examples indicate a need to generalize Shephard's attainability assumption. Note that, as shown in Figure 1, it is possible to rescale both inputs and outputs from point B toward the production frontier (abcde). This suggests considering attainability when both inputs and outputs are rescaled in some way. This idea has been explored by Färe et al. (1985, chapter 5; 1994a, chapter 8) in their proposed hyperbolic measures of technical efficiency. Alternative non-radial measures have also been proposed by Russell, Färe et al. (1985; 1994), Luenberger (1992b, 1995) and Briec.

In this paper, we propose a measure which involves the following additional assumption:

A5. The set A is non-empty, where A is defined as:

$$A = \{\alpha: 0 \leqq \alpha \leqq 1; \exists\ \delta > 0 \text{ such that } (\delta^{1-\alpha} x) \in V(\delta^{-\alpha} y, T)\},$$

for $x \geq 0$, $y \geq 0$.

Assumption A5 is a weak form of Shephard's attainability assumption. To see this, it is sufficient to note that Shephard's input attainability means that $0 \in A$, while Shephard's output attainability implies that $1 \in A$ (see Shephard, chapter 9). Obviously, there are situations where Shephard's attainability axioms are violated (i.e., $0 \notin A$ or $1 \notin A$), yet the set A is non-empty. For example, in figure 1, $0 \notin A$ and $1 \notin A$ at point B, but A is non-empty since it is possible to rescale both inputs and outputs from point B (e.g., toward point c) to generate a resource mix that is technically feasible.

Given the rescaling of both inputs and outputs under assumption A5, we propose the following generalization of Shephard's distance functions.[7]

Definition 2: Define a generalized distance function as:

$$D(x, y, T, \alpha) = \text{Min}_\delta \{\delta: (\delta^{1-\alpha} x) \in V(\delta^{-\alpha} y, T), \delta > 0\}, \tag{2}$$

for $x \geq 0$, $y \geq 0$, and $\alpha \in A$.

Note that the function $1/D(x, y, T, \alpha)$ in (2) becomes Shephard's input distance function when $\alpha = 0$. Alternatively, $D(x, y, T, \alpha)$ becomes Shephard's output distance function when $\alpha = 1$. These two polar cases (i.e., $\alpha = 0$ and $\alpha = 1$) have been investigated in detail by Shephard. The definition in (2) also includes as special cases hyperbolic measures proposed and investigated by Färe et al. (1985 chapter 5; 1994a, chapter 8) and Briec. For example, in the context of (2), the "Farrell graph measure" proposed by Färe et al.'s (1985, p. 110) is equal to $D(x, y, T, 0.5)^2$, with $\alpha = 0.5$. And the "Farrell equiproportionate distance" measure proposed by Briec is equal to $[(D^{-\alpha} - D^{1-\alpha})/2]$, when α is chosen to rescale x and y in the same directions as specified in Briec.[8] Expression (2) is also related to the "generalized Farrell graph measure" proposed by Färe et al. (1985, p. 126). More specifically, compared to D in (2), Färe et al.'s "generalized Farrell graph measure" equals $\text{Min}_\alpha [(D^{1-\alpha} + D^{\alpha})/2]$. Thus, with $0 \leqq \alpha \leqq 1$, our generalized distance function D in (2) extends Shephard's analysis as well as some efficiency measures proposed by Färe et al. and Briec. As such, it nests many cases found in previous literature.[9] Also, as we will see below, our approach provides a convenient and intuitive basis for the analysis of production efficiency.

It is of interest to investigate the properties of the distance function $D(x, y, T, \alpha)$ in (2). These properties are presented in the following propositions. (See the proof in the Appendix).

Proposition 1: For $x \geq 0$, $y \geq 0$ and $\alpha \in A$, the function $D(x, y, T, \alpha)$ in (2) satisfies

$$D(x, y, T, \alpha) \leqq 1 \text{ if and only if } x \in V(y, T) .$$

Proposition 2: For $x \geq 0$, $y \geq 0$ and $\alpha \in A$, the function $D(x, y, T, \alpha)$ in (2) is:

1. almost homogeneous of degree (α-1), α, and 1 in x and y.[10]
2. nonincreasing in x.
3. nondecreasing in y.
4. lower semi-continuous in x and y.
5. independent of α under CRTS,

   decreasing (increasing) in α under IRTS (DRTS) for $x \in V(y, T)$.

Since the function $D(x, y, T, \alpha)$ is a generalization of Shephard's distance functions, it follows that the properties stated in propositions 1 and 2 (with $0 \leqq \alpha \leqq 1$) provide a generalization of earlier results found in the literature. In particular, Proposition 1 shows that $D(x, y, T, \alpha) = 1$ is a transformation function representing the frontier technology for any α, $0 \leqq \alpha \leqq 1$. Also, Proposition 2 establishes how the parameter α influences the generalized distance function $D(\cdot)$ in (2). It shows that $D(\cdot)$ is independent of α under CRTS. In this case, the choice of α is therefore inconsequential: any $\alpha \in A$ would generate the same distance function $D(\cdot)$ in (2). Proposition 2 also indicates how α influences $D(\cdot)$ when the technology departs from CRTS. It shows that $D(\cdot)$ in (2) is decreasing (increasing) in α under IRTS (DRTS) when $x \in V(y, T)$. Thus, under VRTS, the choice of α matters. This raises the question: under VRTS, how to choose α in the definition of the generalized distance function $D(\cdot)$ in (2)?

Here, we suggest choosing the value of $\alpha \in A$ so as to rescale inputs and outputs in a direction that is closest to the frontier technology. Indeed, it seems intuitive that the technical efficiency of a point (x, y) should be evaluated compared to a point on the frontier technology that is in its closest neighborhood. This will be further motivated below in our discussion of technical efficiency. This suggests choosing α as follows:

$$\alpha \in A^* = \text{argmin}_\alpha \{|D(x, y, \alpha, T) - 1|: \alpha \in A\}. \tag{3}$$

Note that, under CRTS, $A^* = A = [0, 1]$ and $D(\alpha)$ is independent of α. In this case, from (3), any choice of α between 0 and 1 appears appropriate (since it has no effect on D). This is consistent with the approach found in previous literature, where under CRTS investigators can choose either $\alpha = 0$ (input-based distance function) or $\alpha = 1$ (output-based distance function) (e.g., Caves et al., 1982b; Färe at al., 1985).

However, under variable return to scale (VRTS), in general $A^* \subseteq [0, 1]$. In such situation, we propose choosing $\alpha \in A^*$ as given in equation (3). This provides guidance in the choice of α in (2) when the technology departs from CRTS. To illustrate, consider the situation where $D(x, y, \alpha, T) \leqq 1$ and $A = \{\alpha: \alpha_L \leqq \alpha \leqq \alpha_M\}$, i.e., where α is bounded between $\alpha_L \geqq 0$ and $\alpha_M \leqq 1$. Then, from Proposition 2 (part 5), equation (3) would imply choosing $\alpha = \alpha_M$ if the technology exhibits DRTS, and $\alpha = \alpha_L$ if the technology exhibits IRTS. This simply reflects the fact that, under DRTS, one can move proportionally faster toward the frontier technology by rescaling outputs. And, under IRTS, one moves proportionally faster toward the frontier by rescaling inputs. Such results provide useful guidelines for the choice of α in the empirical use of the generalized distance function (2).[11]

Finally, note that our proposed choice of α in (3) differs from the one used by Färe et al. (1985, p. 126) in their "Farrell generalized graph measure". As indicated above, in our notation, their measure is $\text{Min}_\alpha [(D^{1-\alpha} + D^\alpha)/2]$. When $D \geq 1$, this is consistent with (3). However, when $D < 1$, their proposed measure involves choosing a scaling direction α that compares (x, y) with a point that is "as far away as possible" on the technology frontier. This seems unappealing and counterintuitive. This unattractive characteristic may help explain why the "Farrell generalized graph measure" has apparently not generated much (if any) empirical work related to efficiency analysis over the last decade.

3- THE MEASUREMENT OF TECHNICAL EFFICIENCY AND PRODUCTIVITY

The distance function D(x, y, T, α) is of special interest in the measurement of production efficiency and productivity. The analysis of economic efficiency has typically centered on the technical, allocative and scale efficiency of production decisions (e.g., Farrell; Färe et al., 1985, 1994a).

3.1- Technical Efficiency

The concept of technical efficiency relates to the question of whether a firm uses the best available technology in its production process. Assume that the firm is observed using inputs-outputs ($x \geq 0$, $y \geq 0$). Following the work of Debreu, Farrell, Farrell and Fieldhouse, and Färe et al. (1985, 1994a), technical efficiency has been defined as the proportional rescaling of inputs or outputs that would bring the firm to the production frontier. This suggests using the generalized distance function (2) as an index of technical efficiency.

Definition 3: Define the technical efficiency index TE as:

$$TE(x, y, T, \alpha) = D(x, y, T, \alpha). \tag{4}$$

where D(x, y, T, α) is given in (2).

When $\alpha = 0 \in A$, the index TE in (4) becomes the input-based Farrell measure of technical efficiency measuring the minimal proportion by which the input vector x can be rescaled while still producing outputs y. Alternatively, when $\alpha = 1 \in A$, the index TE becomes the output-based measure of technical efficiency discussed in Färe et al. (1985, chapter 4; 1994a). And when $0 < \alpha < 1$, equation (4) provides a generalized measure of technical efficiency, rescaling both inputs and outputs toward the frontier technology. As discussed above, this can be of particular interest in situations that do not satisfy Shephard's attainability axiom on the input and/or output side (i.e., when $0 \notin A$ and/or $1 \notin A$). Finally, besides the Farrell measures, note that equation (4) includes as special cases several models found in the literature. The "Farrell graph hyperbolic measure" of technical efficiency proposed by Färe et al. (1985, 1994a) can be seen as a special case: it equals $TE(x, y, T, 0.5)^2$, where $\alpha = 0.5$. Also, our index TE is related to Briec's "Farrell equiproportional distance", which equals $[TE(x, y, T, \alpha)^{-\alpha} - TE(x, y, T, \alpha)^{1-\alpha}]/2$ given an appropriate choice for α (as discussed in footnote 8).

From proposition 1 and equation (4), it follows that $TE(x, y, T, \alpha) \leqq 1$ if and only if $x \in V(y, T)$. This shows that, for a firm choosing (x, y), the technical efficiency index (4) is at most equal to one if the firm uses the technology T. More specifically, TE = 1 if the firm is on the frontier technology T and is said to be technologically efficient. TE < 1 if the firm is not technically efficient. Finally, if TE > 1, then the firm is super efficient and produces beyond the frontier technology. This would be the case if technological progress has taken place and the technology T represents an old technology (see below).

From (4), the technical efficiency index $TE(x, y, T, \alpha)$ inherits all the properties of the distance function $D(x, y, T, \alpha)$ presented in propositions 1 and 2. From proposition 2 (part 5), TE is independent of α under CRTS. This corresponds to the well known result that, under CRTS, the input-based measure of technical efficiency (corresponding to $\alpha = 0$) is equal to the output-based measure of technical efficiency (corresponding to $\alpha = 1$) (e.g., Färe et al., 1985, p. 132). Our analysis shows that, under CRTS, this result generalizes to any value of $\alpha \in A$. In other words, under CRTS, the choice of α in the measurement of TE is inconsequential. Proposition 2 (part 5) also shows that TE is decreasing (increasing) in α under IRTS (DRTS) when $x \in V(y, T)$. This establishes how the parameter α influences the TE index when the technology departs from CRTS. In this case, the choice of α in general affects the measurement of technical efficiency. In section 2, we proposed choosing α according to equation (3). Given the definition of TE in (4), $\alpha \in A^*$ in (3) is chosen so as to generate the technical efficiency index TE that is as close to 1 as possible. This particular choice appears particularly desirable in the sense that, for $x \in V(y, T)$, it provides an upper bound measure of TE which avoids any overstatement of technical inefficiency. For example, given $x \in V(y, T)$, this corresponds to choosing α to be large under DRTS, and to be small under

IRTS. Such a choice of α identifies the shortest path for rescaling inputs-outputs toward the frontier technology.

3.2- <u>Productivity</u>

The TE index in (3) can also be used in the measurement of productivity when the observed input-output vector ($x \geq 0$, $y \geq 0$) is assumed to be on the frontier technology, but this frontier technology is shifting. To see this, consider two situations i = 0, 1. In each situation i, let (x^i, y^i) be the netput vector and let T^i be the associated technology, i = 0, 1. Define the following productivity index

$$PI = D(x^1, y^1, T^0, \alpha)/D(x^1, y^1, T^1, \alpha).$$

PI is the ratio of two distance functions evaluated at the same point (x^1, y^1), but under the different technologies T^0 and T^1. The index PI measures productivity in the sense that PI > 1 for all (x^1, y^1) means that $T^0 \subseteq T^1$, i.e. that the feasible set has expanded between i = 0 and i = 1. When i denotes time, this means that technological progress has taken place between time i = 0 and i = 1. Then, (PI - 1) can be interpreted as measuring the rate of technological change. A related index considered by Caves et al. (1982a, 1982b) is

$$PI' = D(x^1, y^1, T^0, \alpha)/D(x^0, y^0, T^0, \alpha).$$

It is also the ratio of two distance functions D evaluated under the same technology T^0, but at different points (x^0, y^0) and (x^1, y^1). In general, PI an PI' are not equivalent. The following relationship exists between the two indexes

$$PI' = PI \cdot [D(x^1, y^1, T^1, \alpha)/D(x^0, y^0, T^0, \alpha)].$$

Interpreting the term $[D(x^1, y^1, T^1, \alpha)/D(x^0, y^0, T^0, \alpha)]$ as a measure of technical efficiency change, this shows that PI' captures both technical change (through PI) and change in technical efficiency. It also shows that the two productivity indexes PI and PI' become identical <u>under technical efficiency</u>. Indeed, when the observed input-output vectors (x, y) are always on the frontier technology, proposition 1 implies that $D(x^i, y^i, T^i, \alpha) = 1$ for $i = 0, 1$. This implies that PI = PI' under technical efficiency. In this case, the TE index in (4) can also be interpreted as a productivity measure reflecting technological change of the observations (x^1, y^1) compared to the technology T^0. Indeed, $TE(x^1, y^1, T^0, \alpha) = 1$ means that (x^1, y^1) is on the frontier technology T^0. And finding $TE(x^1, y^1, T^0, \alpha) > (<) 1$ implies that (x^1, y^1) is on a higher (lower) productivity level compared to the reference technology T^0. In addition, $TE(x^1, y^1, T^0, \alpha)$ becomes the output-based productivity index investigated by Caves et al. when $\alpha = 1$, and their input-based productivity index when $\alpha = 0$ (Caves et al., 1982a, 1982b). Indeed, given $\alpha = 1$, $TE(x^1, y^1, T^0, 1)$ measures the proportional rescaling of output y toward the frontier iso-product corresponding to the reference technology T^0 (see Caves et al., 1982b, p. 1403). And given $\alpha = 0$, $TE(x^1, y^1, T^0, 0)$ measures the proportional rescaling of inputs x^1 toward the frontier isoquant corresponding to the reference technology T^0 (see Caves et al., 1984b, p. 1407). When $0 \leqq \alpha \leqq 1$, the $TE(x^1, y^1, T^0, 0)$ index in (4) is thus a generalization of the productivity indexes discussed by Caves et al. (1982a, 1982b). As discussed above, this can be of particular interest in productivity analysis when Shephard's attainability assumption is not satisfied (i.e., when $0 \notin A$ and $1 \notin A$), and/or when the technology departs from CRTS (e.g., the analysis conducted by Ray and Desli).

4- <u>THE MEASUREMENT OF ALLOCATIVE EFFICIENCY</u>

In this section, we evaluate the economic performance of a competitive firm that is observed choosing input $x \geq 0$ and output $y \geq 0$ under technology T. This evaluation can be made on the input side (using a cost function), on the output side (using a revenue function), or on both input and output sides (using a profit function).

4.1- <u>Cost-Based Allocative Efficiency</u>

Following Farrell, the concept of allocative efficiency can be related to the ability of the firm to choose its inputs in a cost minimizing way. It reflects whether a technically efficient firm produces at the lowest possible cost, as given by the minimization problem[12]

$$C(r, y, T) = \mathrm{Min}_x \{r'x: x \in V(y, T)\}, \tag{5}$$

where $y \geq 0$, and r is a $(n\times1)$ vector of input prices assumed to be strictly positive $(r > 0)$. The cost function $C(r, y, T)$ is positive linearly homogeneous and concave in r. Given $TE(x, y, T, \alpha)$ in (4), it follows from (2) that the point $(TE^{1-\alpha}\, x, TE^{-\alpha}\, y)$ is technically feasible. This point is a feasible solution to the cost minimization problem (5), although it may not be its optimal solution. If the output vector is evaluated at $[TE(x, y, T, \alpha)^{-\alpha}\, y]$, the cost minimization problem (5) thus implies the following result.

Proposition 3: $C[r, TE(x, y, T, \alpha)^{-\alpha}\, y, T] \leqq r'(TE(x, y, T, \alpha)^{1-\alpha}\, x)$,

for $\alpha \in A$, $x \geq 0$, and $y \geq 0$.

Proposition 3 suggests the following index of allocative efficiency.

Definition 4: Define the cost-based allocative efficiency index AE_C as:

$$AE_C(r, x, y, T, \alpha) = \frac{C(r, TE^{-\alpha}\, y, T)}{r'(TE^{1-\alpha}\, x)} \leqq 1\ , \tag{6}$$

where $TE = TE(x, y, T, \alpha)$.[13] The index $AE_C(r, x, y, T, \alpha)$ in (6) is homogeneous of degree zero in r. It is bounded between zero and one. It is equal to one if the firm is allocatively efficient in the sense of minimizing the cost of producing outputs $(TE^{-\alpha}\, y)$. Alternatively, $AE_C < 1$ implies that the firm is not allocatively efficient. In this case, from (6), $(1 - AE_C)$ measures the extent of allocative inefficiency of the firm: it is the percentage reduction in production cost that the firm can obtain by behaving in a cost minimizing way.

4.2- Revenue -Based Allocative Efficiency

Allocative efficiency can also relate to the ability of the firm to choose its outputs in a revenue maximizing way. This corresponds to the following optimization problem[14]

$$R(p, x, T) = \mathrm{Max}_y \{p'y: x \in V(y, T)\} \tag{7}$$

where $x \geq 0$, and p is $(m\times1)$ vector of output prices assumed to be strictly positive $(p > 0)$. The revenue function $R(p, x, T)$ is linearly homogeneous and convex in p. Note that the point $(TE^{1-\alpha}\, x, TE^{-\alpha}\, y)$ is technically feasible from (2), implying that it is a feasible but not necessarily optimal solution to the revenue maximization problem (7). If the input vector is evaluated at $[TE(x, y, T, \alpha)^{1-\alpha}\, x]$, the maximization problem (7) thus implies the following result.

Proposition 4: $R(p, TE^{1-\alpha}\, x, T) \geqq p'(TE^{-\alpha}\, y)$

for $\alpha \in A$, $x \geq 0$, and $y \geq 0$.

Proposition 4 suggests the following index of allocative efficiency.

Definition 5: Define the revenue-based allocative efficiency index AE_R as

$$AE_R(p, x, y, T, \alpha) = \frac{p'(TE^{-\alpha}\, y)}{R(p, TE^{1-\alpha}\, x, T)} \leqq 1 . \tag{8}$$

The index $AE_R(p, x, y, T, \alpha)$ in (8) is homogeneous of degree zero in p. It is bounded between zero and one. It is equal to one if the firm is allocatively efficient in the sense of maximizing revenue given inputs $(TE^{1-\alpha}\, x)$. Alternatively, $AE_R < 1$ implies that the firm is not allocatively efficient. In this case, $(1- AE_R)$ measures the extent of allocative inefficiency: it is the percentage increase in revenue that the firm can obtain by behaving in a revenue maximizing way.

4.3- Profit-Based Allocative Efficiency

Finally, allocative efficiency can be evaluated in terms of the ability of the competitive firm to maximize profit. This corresponds to the following optimization problem[15]

$$\begin{aligned} \pi(r, p, T) &= Max_{x,y}\, \{p'y - r'x: x \in V(y, T)\} \\ &= Max_y\, \{p'y - C(r, y, T)\} \\ &= Max_x\, \{R(p, x, T) - r'x\}, \end{aligned} \tag{9}$$

where $r > 0$ is the vector of input prices, $p > 0$ is the vector of output prices, and the cost and revenue functions C(·) and R(·) are defined in (5) and (7). The profit function $\pi(r, p, T)$ is linearly homogeneous and convex in (r, p), non-increasing in r, and non-decreasing in p. Note that the point $(TE^{1-\alpha}\, x, TE^{-\alpha}\, y)$ is technically feasible from (2), implying that it is a feasible but not necessarily optimal solution to the profit maximization problem (9). Using (6) and (8), the maximization problem (9) then generates the following results.

<u>Proposition 5</u>: $$\pi(r, p, T) \geqq p'(TE^{-\alpha}\, y) - AE_C\, r'(TE^{1-\alpha}\, x) \geqq p'TE^{-\alpha}\, y - r'(TE^{1-\alpha}\, x) \qquad (10a)$$

$$\pi(r, p, T) \geqq (AE_R)^{-1}\, p'(TE^{-\alpha}\, y) - r'(TE^{1-\alpha}\, x) \geqq p'(TE^{-\alpha}\, y) - r'(TE^{1-\alpha}\, x) \qquad (10b)$$

for $\alpha \in A$, $x \geq 0$, and $y \geq 0$.

Proposition 5 suggests the following index of allocative efficiency.

<u>Definition 6</u>: Define the profit-based allocative efficiency index AE_π as the implicit solution $AE_\pi(r, p, x, y, T, \alpha, \beta)$ to the equation:

$$\pi(r, p, T) = (AE_\pi)^{-\beta}\, p'(TE^{-\alpha}\, y) - (AE_\pi)^{1-\beta}\, r'(TE^{1-\alpha}\, x) \qquad (11)$$

$\alpha \in A$, $x \geq 0$, and $y \geq 0$, for some scalar β, $0 \leqq \beta \leqq 1$.

The index $AE_\pi(r, p, x, y, T, \alpha, \beta)$ defined in (11) is homogeneous of degree zero in (r, p). It is bounded between zero and one: $0 < AE_\pi(r, p, x, y, T, \alpha, \beta) \leqq 1$. The parameter β in (11) can be chosen between 0 (corresponding to downside cost rescaling) and 1 (corresponding to upside revenue rescaling), with $0 < \beta < 1$ reflecting both cost and revenue rescaling. The index AE_π in (11) is equal to one if the firm is allocatively efficient in the sense of maximizing profit. Alternatively, $AE_\pi < 1$ implies that the firm is not allocatively efficient. In this case, the departure of AE_π from one reflects the upside rescaling of revenue (if $\beta = 1$), the downside rescaling of cost (if $\beta = 0$), or both (if $0 < \beta < 1$), that the firm must achieve in order to reach its maximal profit.

What relationships exist between the cost-based index AE_C in (6), the revenue-based index EA_R in (8), and the profit-based index EA_π in (11)? Such relationships can be obtained simply from combining

(11) with (10a) and (10b), yielding the following results.

Proposition 6: $AE_\pi(r, p, x, y, T, \alpha, 0) \leqq AE_C(r, TE^{-\alpha} y, T) \leqq 1$ (12a)

$AE_\pi(r, p, x, y, T, \alpha, 1) \leqq AE_R(p, TE^{1-\alpha} x, T) \leqq 1$ (12b)

for $\alpha \in A$, $x \geq 0$, and $y \geq 0$.

Proposition 6 states that AE_π is a lower-bound for AE_C when $\beta = 0$, and a lower bound for AE_R when $\beta = 1$. These results are intuitive. They simply reflect the fact that, by evaluating both inputs and outputs, profit maximization can uncover more allocative inefficiency than cost minimization (that evaluates only inputs) or revenue maximization (that evaluates only outputs).

5- THE MEASUREMENT OF SCALE EFFICIENCY

Scale efficiency[16] of the firm can be motivated from free entry and exit conditions in the industry, and its implications for long run equilibrium (e.g., see Baumol et al.). It is closely linked with zero profit, which is a necessary condition for long run equilibrium under free entry and exit. Indeed, long run equilibrium is typically defined as a situation where there is no incentive for entry or exit in the industry. Clearly, under free entry, any positive profit provides an incentive for firms to enter the industry. And, under free exit, negative profit for the firm provides an incentive for it to exit the industry. Thus, in the absence of barriers to entry or exit, there is no incentive for entry or exit in the industry only if firm profit is zero.

Again, the evaluation of scale efficiency can be made on the input side (using the cost function), on the output side (using the revenue function), or both input and output sides (using the profit function).

5.1- Cost-Based Scale Efficiency

The efficiency indexes TE and AE_C in (4) and (6) are conditional on outputs y. Thus, they can be interpreted as being conditional on scale y. Yet, the choice of y involves efficiency considerations as well. Whether a firm is producing at an "optimal scale" y can be analyzed through the measurement of returns to scale. In equations (1a)-(1c), we have characterized the nature of returns to scale for the underlying technology T. Alternatively, returns to scale can be analyzed using ray-average cost[17]

$$\{(r'x)/\lambda\colon x \in V(\lambda\, y, T), \lambda > 0, y \geq 0\},$$

which measures the cost per unit of the outputs-scaling factor λ, a positive scalar. Define the following ray-average cost function[18]

$$\begin{aligned} RAC(r, y, T) &= \mathrm{Inf}_{x,\lambda}\, \{(r'x)/\lambda\colon x \in V(\lambda\, y, T), \lambda > 0\} \\ &= \mathrm{Inf}_{\lambda}\, \{C(r, \lambda\, y, T)/\lambda\colon \lambda > 0\}, \end{aligned} \tag{13}$$

where $y \geq 0$, and C(r, y, T) is the cost function defined in (5). The function RAC(r, y, T) in (13) is linearly homogeneous in r and concave in r, and linearly homogeneous in y. It gives the smallest cost per unit of the outputs-scaling factor λ. Under a linearly homogeneous -- CRTS -- technology, the ray-average cost function $[C(r, \lambda\, y, T)/\lambda]$ is independent of λ: a proportional change in outputs yields the same proportional change in production cost. And IRTS (DRTS) corresponds to this ray-average cost being decreasing (increasing) in λ, where a proportional increase in outputs leads to a less than (more than) proportional increase in cost.[19] In the case where the function $[C(r, \lambda\, y, T)/\lambda]$ has a U-shape with respect to λ, then CRTS is attained (locally) at the minimum of the ray-average cost.

Note that $C(r, \lambda\, y, T)/\lambda$ is equal to $C(r, TE^{-\alpha}\, y, T)/TE^{-\alpha}$ when $\lambda = TE^{-\alpha} > 0$. This is a feasible, although not necessarily optimal point in the minimization problem (13). This implies $RAC(r, y, T) \leqq C(r, TE^{-\alpha}\, y, T)/TE^{-\alpha}$. Using proposition 3, this yields the following results.

Proposition 7: $RAC(r, y, T) \leqq C(r, TE^{-\alpha}\, y, T)/TE^{-\alpha} \leqq r'(TE\, x)$,

for $TE^{-\alpha} > 0$, $x \geq 0$, and $y \geq 0$.

Proposition 7 means that the ray average cost is bounded from above by $[C(r, TE^{-\alpha}\, y, T)/TE^{-\alpha}]$. This suggests the following definition of scale efficiency.

Definition 7: Define the cost-based scale efficiency index SE_C as

$$SE_C(r, x, y, T, \alpha) = \frac{RAC(r, y, T)}{C[r, TE(x, y, T, \alpha)^{-\alpha} y, T]/TE(x, y, T, \alpha)^{-\alpha}} \leqq 1. \qquad (14)$$

The scale efficiency index in (14) is bounded between 0 and 1. Values of the vector y satisfying $SE_C = 1$ identify an efficient scale of operation corresponding to the smallest ray-average cost. Alternatively, finding $SE_C < 1$ implies that the vector y is not an efficient scale of operation. In this case, $(1 - SE_C)$ can be interpreted as the maximal relative decrease in the ray-average cost $[C(r, \lambda\, TE^{-\alpha}\, y, T)/(\lambda\, TE^{-\alpha})]$ that can be achieved by proportionally rescaling all outputs toward the efficient scale (where the output vector exhibits <u>locally</u> CRTS).

So far, we have assumed that T represents a variable-return-to-scale (VRTS) technology. Here, we are interested in identifying the production region exhibiting (locally) CRTS. For this purpose, define the constant-return-to-scale (CRTS) technology:

$$T_c = \{(x, y): (\lambda\, x, \lambda\, y) \in T, \text{ for some } \lambda > 0\}, \qquad (15)$$

The cone technology T_c generated by T is the smallest CRTS technology that contains T. It satisfies $T \subseteq T_c$. Note that, using (15), the RAC function (13) can be alternatively expressed as:

$$\begin{aligned} RAC(r, y, T) &= \text{Inf}_{x,\lambda}\ \{(r'x)/\lambda: x \in V(\lambda\, y, T), \lambda > 0, y \geq 0\} \\ &= \text{Inf}_{X,\lambda}\ \{r'X: \lambda\, X \in V(\lambda\, y, T), \lambda > 0, y \geq 0\}, \text{ where } X = x/\lambda, \\ &= \text{Inf}_{X}\ \{r'X: X \in V(y, T_c), y \geq 0\} \\ &= C(r, y, T_c). \end{aligned}$$

This implies that the scale efficiency index SE_C in (14) can be written as

$$SE_C(r, x, y, T, \alpha) = \frac{C(r, y, T_c)}{C[r, TE(x, y, T, \alpha)^{-\alpha} y, T]/TE(x, y, T, \alpha)^{-\alpha}} \leqq 1. \qquad (14')$$

Equation (14') provides a convenient alternative measure which can prove useful in the analysis of scale efficiency.

5.2- Revenue-Based Scale Efficiency

The cost-based scale efficiency index SE_C in (14) is conditional on scale y. This implicitly measures firm size by its outputs y. As an alternative, we now consider the case where firm size is measured by its inputs x. As in the context of the efficiency indexes TE and AE_R in (4) and (8), this requires measurements that are conditional on x. Whether the firm is producing at an "optimal scale" x can be analyzed using ray-average revenue[20]

$$\{(p'y)/\lambda: (\lambda\, x) \in V(y, T), \lambda > 0, y \geq 0\},$$

which measures the revenue per unit of the inputs-scaling factor λ, a positive scalar. Define the following ray-average revenue function[21]

$$\begin{aligned} RAR(p, x, T) &= \text{Sup}_{y,\lambda}\ \{p'y/\lambda: (\lambda\, x) \in V(y, T), \lambda > 0\} \\ &= \text{Sup}_{\lambda}\ \{R(p, \lambda\, x, T)/\lambda, \lambda > 0\} \qquad (16) \end{aligned}$$

where $x \geq 0$, and R(p, x, T) is the revenue function defined in (7). The function RAR(p, x, T) in (16) is linearly homogeneous and convex in p, and linearly homogeneous in x. It gives the largest revenue per unit of the inputs-scaling factor λ. In this context, under a linearly homogeneous -- CRTS -- technology, the ray-average revenue function $[R(p, \lambda\, x, T)/\lambda]$ is independent of λ: a proportional change in inputs yields the same proportional change in revenue. And IRTS (DRTS) corresponds to this ray-average revenue function being increasing (decreasing) in λ, where a proportional increase in inputs leads to a more than (less than) proportional increase in revenue. In the case where the function $[R(p, \lambda\, x, T)/\lambda]$ has an inverted-U shape, then CRTS is attained (locally) at the maximum of the ray-average revenue.

Note that $R(r, \lambda\, x, T)/\lambda$ is equal to $R(r, TE^{1-\alpha}\, x, T)/TE^{1-\alpha}$ when $\lambda = TE^{1-\alpha}$. This is a feasible, although not necessarily optimal point in the maximization problem (16). This implies $RAR(p, x, T) \geqq R(p, TE^{1-\alpha}\, x, T)/TE^{1-\alpha}$. Using proposition 4, this yields the following results.

Proposition 8: $RAR(p, x, T) \geqq R(p, TE^{1-\alpha}\, x, T)/TE^{1-\alpha} \geqq p'(TE^{-1}\, y)$

for $TE^{1-\alpha} > 0$, $x \geq 0$, and $y \geq 0$.

Proposition 8 means that the ray-average revenue is bounded from below by $[R(p, TE^{1-\alpha}\, x, T)/TE^{1-\alpha}]$. This suggests the following definition of revenue-based scale efficiency.

Definition 8: Define the revenue-based scale efficiency index SE_R as

$$SE_R(p, x, y, T, \alpha) = \frac{R(p, TE^{1-\alpha}\, x, T)/TE^{1-\alpha}}{RAR(p, x, T)} \leqq 1. \tag{17}$$

The scale efficiency index $SE_R(p, x, y, T, \alpha)$ in (17) is homogeneous of degree zero in p. It is bounded between zero and one. Values of the vector x satisfying $SE_R = 1$ identify an efficient scale of operation corresponding to the largest ray-average revenue. Alternatively, finding $SE_R < 1$ implies that the vector x is an inefficient scale of operation. In this case, $(1 - SE_R)$ can be interpreted as the maximal relative increase in the ray-average revenue $R(p, \lambda\, TE^{1-\alpha}\, x, T)/(\lambda\, TE^{1-\alpha})$ that can be achieved by proportional rescaling all inputs toward the efficient scale (where the input vector exhibits <u>locally</u> CRTS).

Note that, using the CRTS technology T_c defined in (15), the RAR function (16) can be alternatively expressed as

$$\begin{aligned} RAR(p, x, T) &= \mathrm{Sup}_{y,\lambda}\, \{p'y/\lambda: (\lambda\, x) \in V(y, T), \lambda > 0\} \\ &= \mathrm{Sup}_{Y,\lambda}\, \{p'Y: (\lambda\, x) \in V(\lambda\, Y, T), \lambda > 0\}, \text{ where } Y = y/\lambda, \\ &= \mathrm{Sup}_{Y}\, \{p'Y: x \in V(Y, T_c)\} \\ &= R(p, x, T_c) \end{aligned}$$

for $x \geq 0$. This implies that the scale efficiency index SE_R in (17) can be alternatively written as

$$SE_R(p, x, y, T, \alpha) = \frac{R(p, TE^{1-\alpha}\, x, T)/TE^{1-\alpha}}{R(p, x, T_c)} \leqq 1. \tag{17'}$$

Equation (17') provides a convenient alternative measure which can prove useful in scale efficiency analysis.

5.3- Profit-Based Scale Efficiency

Finally, scale efficiency can be evaluated in terms of both inputs and output. This can be done using the cost-to-revenue ratio[22]

$\{r'x/(p'y): x \in V(y, T), x \geq 0, y \geq 0\}$,

which measures the cost per unit of revenue. Define the following cost-revenue function[23]

$$CR(r, p, T) = \text{Inf}_{x,y} \{r'x/(p'y), x \in V(y, T), x \geq 0, y \geq 0\} \quad (18)$$

where $r > 0$ and $p > 0$ are input and output price vectors, respectively. The function CR(r, p, T) in (18) is homogeneous of degree zero in (r, p). Note that equation (18) could be equivalently written as

$$CR(r, p, T) = \text{Inf}_{x,y,\lambda} \{\lambda: \lambda^{\beta} p'y - \lambda^{\beta-1} r'x \geqq 0, x \in V(y, T), x \geq 0, y \geq 0, \lambda > 0\}, \quad (18')$$

for any β, $0 \leqq \beta \leqq 1$. When $\beta = 1$, this shows that CR measures the smallest proportional rescaling of output prices p that the firm could sustain without facing a negative profit. Alternatively, when $\beta = 0$, CR measures the largest proportional rescaling of input prices r that the firm could face without generating negative profit. More generally, with $0 \leqq \beta \leqq 1$, CR is a measure of both increase in input prices and/or decrease in output prices that the firm could sustain without obtaining negative profit.

Expression (18') implies that, if CR = 1, then the choices for x and y in (18) are the same as the profit maximizing choices in (9). More generally, (18') means the choices for x and y in (18) are always consistent with profit maximizing behavior when (9) is evaluated at prices $(CR^{\beta-1} r, CR^{\beta} p)$. This provides a formal linkage between the minimization of cost-revenue ratio and profit maximization. Assuming that a profit maximizing solution in (9) exists and is unique, denote by $x^*(r, p, T)$ and $y^*(r, p, T)$ the profit-maximizing inputs and outputs. Let $C^*(r, p, T) = r'x^*(r, p, T)$, and $R^*(r, p, T) = p'y^*(r, p, T)$ be the corresponding profit-maximizing cost and revenue. It follows that

$$CR(r, p, T) = C^*(CR^{\beta-1} r, CR^{\beta} p, T)/R^*(CR^{\beta-1} r, CR^{\beta} p, T) \quad (19)$$

for any β, $0 \leqq \beta \leqq 1$.[24] Clearly, if CR = 1, then $CR(r, p, T) = C^*(r, p, T)/R^*(r, p, T) = 1$, which implies zero profit: $\pi(r, p, T) = R^*(r, p, T) - C^*(r, p, T) = 0$. Thus, finding CR(r, p, T) = 1 necessarily implies zero profit, and no incentive for the current firm to exit or for potential firms to enter the industry. In other words, under free entry and exit, CR(r, p, T) = 1 identifies a scale efficient firm in a long run market equilibrium.

This suggests that finding CR(r, p, T) $\neq$ 1 will typically not be associated with scale efficiency. Finding CR(r, p, T) < 1 means that the firm can face a decrease in output prices and/or an increase in input prices without obtaining negative profit. Since the profit function $\pi(r, p, T)$ is non-decreasing in p and non-increasing in r, this means that the profit-maximizing firm exhibits positive profit: $\pi(r, p, T) > 0$. Under free entry, this would give an incentive for potential firms to enter the industry. Thus, it could not be a long run market equilibrium. Alternatively, finding CR(r, p, T) > 1 means that the firm must face an increase in output prices and/or a decrease in input prices in order to avoid negative profit. Again since the profit function $\pi(r, p, T)$ is non-decreasing in p and non-increasing in r, this means that the profit-maximizing firm exhibits negative profit: $\pi(r, p, T) < 0$. Under free entry, this would give an incentive for the current firm to shut down and exit the industry. Again, this could not be a long run market equilibrium.

This establishes the following relationship between CR and profit π.

<u>Proposition 9</u>: The cost-revenue ratio CR(r, p, T) in (18) is less than (equal to, or greater than) one as profit $\pi(r, p, T)$ is greater than (equal to, or less than) zero.

By linking explicitly production decisions for x and y with zero profit, CR = 1 in (18) thus identifies long run market equilibrium conditions under which there is no incentive for entry or exit in the industry. These are precisely the conditions associated with scale efficiency.

Equation (18) defines CR(r, p, T) as the lower bound of the cost-to-revenue ratio r'x/(p'y). Since attaining this lower bound means obtaining the smallest feasible cost given outputs y, and the largest feasible revenue given inputs x, this implies that the choice of (x, y) in (18) is necessarily consistent with cost minimization and revenue maximization. It follows that (18) can be alternatively expressed as

$$CR(r, p, T) = \text{Inf}_y \{C(r, y, T)/p'y: y \geq 0\} \quad (20a)$$

$$= 1/[\text{Sup}_x \{R(p, x, T)/(r'x): x \geq 0\}] \quad (20b)$$

where C(r, y, T) and R(p, x, T) are the cost and revenue function defined in (5) and (7), respectively.

It is useful to compare equation (20a) with equation (13). It is clear that the minimization problems in (13) and (20a) are similar, except that the former is more restrictive in the sense that it only allows a proportional rescaling of y. This gives

$$CR(r, p, T) = \text{Inf}_y \{C(r, y, T)/p'y: y \geq 0\}$$

$$= \text{Inf}_y \{C(r, p'y \cdot (y/p'y), T)/p'y: y \geq 0\}$$

$$\leqq \text{Inf}_\lambda \{C(r, \lambda \cdot (y/p'y), T)/\lambda: \lambda > 0\}$$

$$= RAC(r, y/p'y, T)$$

$$= RAC(r, y, T)/(p'y)$$

since RAC(r, y, T) is linearly homogeneous in y. Using propositions 3 and 7, this yields the following results.

<u>Proposition 10</u>:

$$CR(r, p, T) \leqq RAC(r, y, T)/(p'y) \leqq C(r, TE^{-\alpha}\, y, T)/(TE^{-\alpha}\, p'y) \leqq TE\; r'x/(p'y)$$

for $x \geq 0$, and $y \geq 0$.

Similarly, compare equation (20b) with equation (16). Note that the maximization problems in (16) and (20b) are similar, except that the former is more restrictive in the sense that it only allows a proportional rescaling of x. Following similar steps as above, this implies that $CR(r, p, T) \leqq r'x/RAR(p, x, T)$. Using propositions 4 and 8, this yields the following results.

<u>Proposition 11</u>:

$$CR(r, p, T) \leqq r'x/RAR(p, x, T) \leqq TE^{1-\alpha}\, r'x/R(p, TE^{1-\alpha}\, x, T) \leqq TE\; r'x/(p'y)$$

for $x \geq 0$, and $y \geq 0$.

Propositions 10 and 11 establish relationships among the cost-revenue function CR(·), the ray-average cost function RAC(·), the cost function C(·), the ray-average revenue function RAR(·), and the revenue function R(·).

Note that the profit-maximizing levels x^* and y^* in (9) are always feasible but not necessarily optimal in the minimization problem (18). This implies the following result.

<u>Proposition 12</u>:

$$CR(r, p, T) \leqq C^*(r, p, T)/R^*(r, p, T),$$

where CR(r, p, T) is given in (18), and $C^*(r, p, T)$ and $R^*(r, p, T)$ are the profit-maximizing cost and revenue, respectively.

Proposition 12 establishes that CR(r, p, T) is a lower bound on the ratio of profit maximizing cost to revenue. It suggests the following definition of profit-based scale efficiency.

<u>Definition 9</u>: Define the profit-based scale efficiency index SE_π as

$$SE_\pi(r, p, T) = \frac{CR(r, p, T)}{C^*(r, p, T)/R^*(r, p, T)} \leq 1. \tag{21}$$

The profit-based scale efficiency index $SE_\pi(r, p, T)$ in (21) is homogeneous of degree zero in (r, p). It is bounded between zero and one. Finding $SE_\pi = 1$ means that the firm is choosing x and y such that it obtains the lowest possible cost-to-revenue ratio. From equation (19), CR(r, p, T) = 1 implies that CR(r, p, T) = $C^*(r, p, T)/R^*(r, p, T)$. Thus, finding CR = 1 necessarily implies that the scale efficiency index SE_π in (21) attains its maximum ($SE_\pi = 1$). Thus CR = 1 identifies both a firm in long run equilibrium (as argued above) and a firm that is scale efficient.

Alternatively, finding $SE_\pi(r, p, T) < 1$ identifies a departure from scale efficiency. In this case, ($1 - SE_\pi$) is a measure of the extent of scale inefficiency. It can be interpreted as the relative change in output-input price ratio that the firm would sustain in its move toward long run equilibrium.

How does the profit-based scale efficiency index SE_π relate to the corresponding indexes SE_C and SE_R defined earlier? Using proposition 10, equations (14) and (21) imply

$$SE_\pi(r, p, T) \leqq RAC(r, y, T)\, R^*(r, p, T)/[p'y\, C^*(r, p, T)]$$

$$= SE_C \cdot [C(r, TE^{-\alpha}\, y, T)/(TE^{-\alpha}\, p'y)]/[C^*(r, p, T)/R^*(r, p, T)].$$

Note that, in the case where the firm maximizes profit, then $TE = 1$, $C(r, y, T) = C^*(r, p, T)$, and $p'y = R^*(r, p, T)$. It follows for a profit-maximizing firm that $SE_\pi \leq SE_C$, i.e. that the profit-based scale efficiency index SE_π is a lower bound on the cost-based scale efficiency index SE_C.

Similarly, using proposition 11, equations (17) and (21) imply

$$SE_\pi(r, p, T) \leqq r'x\, R^*(r, p, T)/[RAR(p, x, T)\, C^*(r, p, T)]$$

$$= SE_R \cdot [TE^{1-\alpha}\, r'x/R(p, TE^{1-\alpha}\, x, T)]/[C^*(r, p, T)/R^*(r, p, T)].$$

Again, in the case where the firm maximizes profit, then $TE = 1$, $r'x = C^*(r, p, T)$, and $R(p, x, T) = R^*(r, p, T)$. It follows for a profit maximizing firm that $SE_\pi \leq SE_R$, i.e. that the profit-based scale efficiency index SE_π is a lower bound on the revenue-based scale efficiency index SE_R.

Finally, what is the relationship between the profit-based scale efficiency index SE_π and the profit-based allocative efficiency index AE_π defined in equation (11)? We have seen that the optimization problem in (18) is always consistent with profit maximization when (9) is evaluated at prices $(CR^{\beta-1}\, r, CR^{\beta}\, p)$. Using the fact that profit is homogeneous of degree one in prices (r, p), note that zero profit is obtained in (18) at prices $(CR^{\beta-1}\, r, CR^{\beta}\, p)$ for <u>any</u> scalar β, $0 \leqq \beta \leqq 1$. Equation (11) evaluated at prices $(CR^{\beta-1}\, r, CR^{\beta}\, p)$ thus implies

$$AE_\pi(CR^{\beta-1}\, r, CR^{\beta}\, p, x, y, T, \alpha, \beta) = CR\, p'y/(TE\, r'x) \tag{22}$$

for <u>any</u> β, $0 \leqq \beta \leqq 1$. Equation (22) suggests an alternative (but related) definition to profit-based allocative efficiency.

Definition 10: Define the profit-based allocative efficiency index AE_π' as

$$AE_\pi'(r, p, x, y, T, \alpha) = \frac{C^*(r, p, T)/R^*(r, p, T)}{r'(TE\ x)/p'y}, \tag{23}$$

where $C^*(r, p, T) = r'x^*(r, p, T)$, $R^*(r, p, T) = p'y^*(r, p, T)$, and $x^*(r, p, T)$ and $y^*(r, p, T)$ are the profit maximizing inputs and outputs in (9).

How does AE_π' in (23) relate to AE_π in (11)? In general, they differ from each other. However, they become identical when they are both evaluated at prices ($CR^{\beta-1}$ r, CR^β p). To see that, evaluating (23) at ($CR^{\beta-1}$ r, CR^β p) and using the implied zero-profit condition gives $AE_\pi'(CR^{\beta-1} r, CR^\beta p, x, y, T, \alpha) = CR\ p'y/(TE\ r'x)$. Combining this with (22) implies that $AE_\pi'(CR^{\beta-1} r, CR^\beta p, x, y, T, \alpha) = AE_\pi(SE_\pi^{\beta-1} r, SE_\pi^\beta p, x, y, T, \alpha, \beta)$ for <u>any</u> β, $0 \leqq \beta \leqq 1$. Thus, when they are both evaluated at prices ($CR^{\beta-1}$ r, CR^β p), the profit-based allocative efficiency index AE_π' in (23) inherits the properties of AE_π in (11) derived above. The relationships between AE_π, AE_C and AE_R were presented in proposition 6. We will argue below that the alternative index AE_π' provides a more convenient basis for combining allocative efficiency with scale efficiency.

6- <u>ECONOMIC INTERPRETATIONS</u>

6.1- <u>Cost-Based Measures</u>

In the investigation of cost-based scale efficiency, we have made use of the ray-average cost $\{(r'x)/\lambda: x \in V(\lambda y, T), \lambda > 0\}$. We show here that all the cost-based efficiency indexes proposed in previous sections can be interpreted as ratios of ray-average costs.

First, note that the technical efficiency index (4) can be written as:

$$TE(x, y, T, \alpha) = \frac{r'(TE^{1-\alpha}\ x)/TE^{-\alpha}}{r'x/1}. \tag{24}$$

Equation (24) is a ratio of two ray-average costs. The numerator is the technically efficient ray-average cost evaluated at inputs ($TE^{1-\alpha}$ x) and at outputs-scaling factor $\lambda = TE^{-\alpha}$. And the denominator is the <u>actual ray-average cost</u> evaluated at x and $\lambda = 1$ (i.e., without rescaling outputs y). In this context, (1 - TE) measures the proportional reduction in ray-average cost that the firm can achieve by becoming technically efficient. A similar interpretation applies if TE is a productivity index evaluating the data point (x, y) compared to the reference technology T.

Second, the cost-based allocative efficiency index (6) can be alternatively written as:

$$AE_C(r, x, y, T, \alpha) = \frac{C(r, TE^{-\alpha} y, T)/TE^{-\alpha}}{r'(TE^{1-\alpha} x)/TE^{-\alpha}}, \tag{25}$$

evaluated at TE = TE(x, y, T, α) > 0. Equation (25) is a ratio of two ray-average costs. The numerator is the technically and allocatively efficient ray-average cost evaluated at cost minimizing inputs and at outputs-scaling factor $\lambda = TE^{-\alpha} > 0$. And the denominator is the technically efficient ray-average cost evaluated at inputs ($TE^{1-\alpha}$ x) and at $\lambda = TE^{-\alpha} > 0$. In this context, (1 - AE_C) measures the proportional reduction in ray-average cost that the firm can achieve by becoming cost-allocatively efficient.

Third, the scale efficiency index (14) can also be interpreted a ratio of two ray-average costs. The numerator is the RAC function defined in (13) and identifying local CRTS (see (14')). And the denominator is the technically and allocatively efficient ray-average cost evaluated at cost minimizing inputs and at outputs-scaling factor $\lambda = TE^{-\alpha} > 0$. Thus, (1 - SE_C) measures the proportional reduction in ray-average cost that the firm can achieve by becoming scale efficient from a cost viewpoint.

Finally, note that these indexes can be combined together. For example, consider the two indexes TE and AE_C. We have seen that they can both be interpreted as being conditional on scale y. From (24) and (25), they can be combined into a cost-based economic efficiency index given scale y, EE_C, where:

$$EE_C = TE \cdot AE_C = \frac{C(r, TE^{-\alpha} y, T)/TE^{-\alpha}}{r'x/1} \leqq 1$$

evaluated at $TE = TE(x, y, T, \alpha) > 0$. The economic efficiency index $EE_C = (TE \cdot AE_C)$ is bounded between 0 and 1. And $(1 - EE_C)$ measures the proportional reduction in ray-average cost that can be achieved by the firm becoming both technically and cost-allocatively efficient.

The scale efficiency index SE_C in (14) can also be combined with the efficiency indexes TE and AE_C. More specifically, from (14), (24) and (25), an overall index of economic efficiency, OE_C, can be defined as the product of the three indexes TE, AE_C and SE_C:

$$OE_C = TE \cdot AE_C \cdot SE_C = \frac{RAC(r, y, T)}{r'x/1} \leqq 1 .$$

Again, the overall efficiency index OE_C is bounded between 0 and 1, and $(1 - OE_C)$ measures the proportional reduction in the ray-average cost $(r'x/1)$ that a firm can achieve by becoming technically, allocatively and scale efficient.

6.2- <u>Revenue-Based Measures</u>

Similar arguments can be presented from the revenue side. Here, the interpretation will focus on ray-average revenue $\{(p'y)/\lambda: (\lambda x) \in V(y, T), \lambda > 0\}$. We show that all the revenue-based efficiency indexes proposed in previous sections can be interpreted as ratios of ray-average revenues. First, note the technical efficiency index (4) can be written as

$$TE(x, y, T, \alpha) = \frac{p'y/1}{p'(TE^{-\alpha} y)/TE^{1-\alpha}}, \tag{26}$$

which is a ratio of the actual and technically efficient ray-average revenues. Second, the revenue-based allocative efficiency index (8) can be written as

$$AE_R(r, x, y, T, \alpha) = \frac{p'\, TE^{-\alpha}\, y/TE^{1-\alpha}}{R(p, TE^{1-\alpha}\, x, T)/TE^{1-\alpha}} \leqq 1, \tag{27}$$

which is a ratio of the technically efficient, and the technically and allocatively efficient ray-average revenues. Third, the revenue-based scale efficiency index (17) is also a ratio of two ray-average revenues. Finally, these indexes can be combined to generate a revenue-based economic efficiency index EE_R and a revenue-based overall efficiency index OE_R

$$EE_R = TE \cdot AE_R = \frac{p'y/1}{R(p, TE^{1-\alpha} x, T)/TE^{1-\alpha}} \leqq 1,$$

and

$$OE_R = TE \cdot AE_R \cdot SE_R = \frac{p'y/1}{RAR(p, x, T)} \leqq 1.$$

The overall revenue-based efficiency index OE_R evaluates jointly the technical, allocative and scale efficiency of the firm using ray-average revenue. And $(1 - OE_R)$ measures the maximal percentage increase in ray-average revenue that the firm can obtain by becoming technically, allocatively and scale efficient.

6.3- <u>Profit-Based Measures</u>

Similar arguments hold for the profit measures. Here, the interpretation will focus on the cost-to-revenue ratio $\{(r'x/(p'y)\colon x \in V(y, T), x \geq 0, y \geq 0\}$. Indeed, all the profit-based efficiency indexes proposed in previous sections can be interpreted as ratios of cost-to-revenue ratios. First, note the technical efficiency index (4) can be written as

$$TE(x, y, T, \alpha) = \frac{r'(TE^{1-\alpha} x)/[p'(TE^{-\alpha} y)]}{r'x/(p'y)} \leqq 1, \tag{28}$$

which is a ratio of the technically efficient and actual cost-to-revenue ratios. Second, from (23), the profit-based allocative efficiency index AE_π'can be written as

$$AE_\pi'(r, p, x, y, T, \alpha) = \frac{C^*(r, p, T)/R^*(r, p, T)}{r'(TE^{1-\alpha} x)/[p'(TE^{-\alpha} y)]} \leqq 1, \tag{29}$$

which is a ratio of "technically and allocatively efficient" and "technically efficient" cost-to-revenue ratios.

Third, we have proposed to characterize scale efficiency using SE_π in (21). Again, the profit-based scale efficiency index (21) is a ratio of cost-revenue ratios.

Finally these indexes can be combined to generate a profit-based economic efficiency index EE_π and a profit-based overall efficiency index OE_π

$$EE_\pi = TE \cdot AE_\pi' = \frac{C^*(r, p, T)/R^*(r, p, T)}{r'x/(p'y)} \leqq 1,$$

and

$$OE_\pi = TE \cdot AE_\pi' \cdot SE_\pi = \frac{CR(r, p, T)}{r'x/(p'y)} \leqq 1.$$

The overall profit-based efficiency index OE_π evaluates jointly the technical, allocative and scale efficiency of the firm using cost-revenue ratios. The measure ($1 - OE_\pi$) gives the largest percentage decrease in cost-revenue ratio that the firm can sustain by becoming technically, allocatively and scale efficient. Alternatively, ($1 - OE_\pi$) measures the relative price change that can be transferred to consumers in the form of lower output prices as the firm moves toward long run market equilibrium. In this case, the benefits from production efficiency are entirely captured by consumers under free entry and exit.

7- CONCLUDING REMARKS

This paper proposes a generalized measure of Shephard's distance functions for a general multi-input multi-output technology. The generalization involves rescaling both input and outputs toward the frontier technology. This allows for a weaker form of attainability than the one assumed by Shephard, thus extending the range of applications of distance functions. Our proposed approach helps settle an issue found in the empirical use of Shephard's distance functions: under variable return to scale (VRTS), the input-distance function and the output-distance function provide different measures of efficiency or productivity. We have shown how our generalized distance function can be used to avoid this dilemma.

Our generalization is found to be useful in the investigation of economic efficiency and productivity analysis. Building on our generalized distance function, indexes are proposed to measure technical, allocative and scale efficiency, as well as productivity. Such indexes can be conveniently interpreted as ratios of ray-average cost, ray-average revenue or cost-to-revenue ratios. Also, the technical, allocative and scale efficiency indexes can be easily combined into overall efficiency indexes, again with simple and intuitive economic interpretations. These indexes provide a convenient basis for the economic investigation of production efficiency and technical change. In particular, given either a parametric representation (e.g., translog) or a nonparametric representation (e.g., as in DEA) of the underlying technology, all the proposed indexes can be estimated as the solution of fairly simple optimization problems. As such, our proposed efficiency and productivity indexes should help refine economic analyses of production efficiency and technical change.

References

Aczél, J. Lectures on Functional Equations and their Applications Academic Press, New-York, 1966.

Afriat, S. "Efficiency Estimates of Production Functions" International Economic Review 13(1972):568-598.

Banker, R.D. "Estimating Most Productive Scale Size Using Data Envelopment Analysis" European Journal of Operational Research 17(1984):35-44.

Banker, R.D., A. Charnes and W.W. Cooper "Models for the Estimation of Technical and Scale Efficiencies in Data Envelopment Analysis" Management Science 30(1984):1078-1092.

Banker, R.D., and A. Maindiratta "Nonparametric Analysis of Technical and Allocative Efficiencies in Production" Econometrica 56(1988):1315-1332.

Bauer, P.W. "Recent Developments in the Econometric Estimation of Frontiers" Journal of Econometrics 46(1990):39-56.

Baumol, William J., J.C. Panzar, and R.D. Willig Contestable Markets and the Theory of Industry Structure Harcourt Brace Jovanovich, Inc., New York, 1982.

Briec, W. "A Graph-Type Extension of Farrell Technical Efficiency Measure" Journal of Productivity Analysis 8(1997): 95-110.

Caves, D.W., L.R. Christensen and W.E. Diewert "Multilateral Comparisons of Output, Input and Productivity Using Superlative Index Numbers" The Economic Journal 92(1982a):73-86.

Caves, D.W., L.R. Christensen and W.E. Diewert "The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity" Econometrica 50(1982b):1393-1414.

Chambers, R.G. Applied Production Analysis: A Dual Approach Cambridge University Press, Cambridge, 1988.

Chambers, R.G., Y. Chung and R. Färe "Benefit and Distance Functions" Journal of Economic Theory 70(1996a):407-419.

Chambers, R.G., Y. Chung and R. Färe "Profit, Directional Distance Functions and Nerlovian Efficiency" Discussion Paper 96-05, Department of Economics, Southern Illinois University, Carbondale, 1996b.

Debreu, G. "The Coefficient of Resource Utilization" Econometrica 19(1951):273-292.

Diewert, W.E. "Exact and Superlative Index Numbers" Journal of Econometrics 4(1976):115-145.

Färe, Rolf, S. Grosskopf, and C.A.K. Lovell The Measurement of Efficiency of Production Kluwer Nijhoff Publishing, Boston, 1985.

Färe, Rolf and Thomas M. Mitchell "On Shephard's Attainability Axiom" Journal of Institutional and Theoretical Economics 143(1987):343-350.

Färe, Rolf, S. Grosskopf, and C.A.K. Lovell "Scale Elasticity and Scale Efficiency" Journal of Institutional and Theoretical Economics 144(1988):721-729.

Färe, Rolf, S. Grosskopf, and C.A.K. Lovell Production Frontiers Cambridge University Press, Cambridge, 1994a.

Färe, Rolf, S. Grosskopf, M. Norris and Z. Zhang "Productivity Growth, Technical Progress and Efficiency Change in Industrialized Countries" American Economic Review 84(1994b):66-83.

Farrell, M.J. "The Measurement of Productive Efficiency" Journal of the Royal Statistical Society, Series A, 120(1957):253-290.

Farrell, M.J. and M. Fieldhouse "Estimating Efficient Production under Increasing Return to Scale" Journal of the Royal Statistical Society Series A, 125(1962):252-267.

Forsund, F.R., C.A.K. Lovell and P. Schmidt "A Survey of Frontier Production Functions and their Relationship to Efficiency Measurement" Journal of Econometrics 13(1980):5-25.

Lau, L.J. "Profit Functions of Technologies with Multiple Inputs and Outputs" The Review of Economic and Statistics 54(1972):281-289.

Luenberger, D.G. "Benefit Functions and Duality" Journal of Mathematical Economics 21(1992a):461-481.

Luenberger, D.G. "New Optimality Principles for Economic Efficiency and Equilibrium" Journal of Optimization Theory and Applications 75(1992b):221-264.

Luenberger, D.G. Microeconomic Theory McGraw-Hill, Boston, 1995.

Malmquist, S. "Index Numbers and Indifference Surfaces" Trabajos de Estatistica 4(1953):209-242.

Ray, S.C. and E. Desli "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries: Comment" American Economic Review 87(1997):1033-1039.

Russell, R.R. "Measures of Technical Efficiency" Journal of Economic Theory 35:1 (1985): 109-126.

Seiford, L.M. and R.M. Thrall "Recent Developments in DEA: The Mathematical Programming Approach to Frontier Analysis" Journal of Econometrics 46(1990):7-38.

Shephard, Ronald W. Theory of Cost and Production Functions Princeton University Press, Princeton, New Jersey, 1970.

Teusch, W. "Aufbau und Gewinnung Shephardscher Produktionsfunktionen unter Berücksichtigung Empirischer Aspekte" Königstein, 1983.

APPENDIX

Proof of Proposition 1: Assume that $D(x, y, T, \alpha) \le 1$. By definition, $\text{Min}_\delta \{\delta: (\delta^{1-\alpha} x) \in V(\delta^{-\alpha} y, T), \delta > 0\} \le 1$. First, consider $\delta = 1$. From assumption A4, the set T is closed, implying that $x \in V(y, T)$. Second, consider $\delta < 1$. If $\alpha = 0$, then $(x\, \delta) \in V(y, T)$ and $(x\, \delta) < x$ imply that $x \in V(y, T)$ from A3. If $0 < \alpha \leqq 1$, then $(\delta^{1-\alpha} x) \in V(\delta^{-\alpha} y, T)$ and $(\delta^{1-\alpha} x) \leqq x$ imply that $x \in V(\delta^{-\alpha} y, T)$ from A3. And from A2, $(\delta^{-\alpha} y) \geqq y$ implies that $x \in V(y, T)$.

Conversely, assume that $x \in V(y, T)$. Then, $(1^{1-\alpha} x) \in V(1^{-\alpha} y, T)$, $0 \leqq \alpha \leqq 1$, implying from (2) that $D(x, y, T, \alpha) \leqq 1$.

Proof of Proposition 2:

1. To show the almost homogeneity property, note that

$$\begin{aligned} D(x, y, T, \alpha) &= \text{Min}_\delta \{\delta: (\delta^{1-\alpha} x) \in V(\delta^{-\alpha} y, T), \delta > 0\} \\ &= \text{Min}_\gamma \{\gamma/\lambda: (\lambda^{\alpha-1} \gamma^{1-\alpha} x) \in V(\lambda^{\alpha} \gamma^{\alpha} y, T), \gamma > 0\},\ \lambda > 0 \\ &= \lambda^{-1} \text{Min}_\gamma \{\gamma: (\lambda^{\alpha-1} \gamma^{1-\alpha} x) \in V(\lambda^{\alpha} \gamma^{\alpha} y, T), \gamma > 0\},\ \lambda > 0 \\ &= \lambda^{-1} D(\lambda^{\alpha-1} x, \lambda^{\alpha} y, T, \alpha),\ \lambda > 0. \end{aligned}$$

2. The function $D(x, y, T, \alpha)$ is decreasing in x because, from assumptions A2 and A3, $x' \geqq x$ implies

$$\text{Min}_\delta \{\delta: (\delta^{1-\alpha} x) \in V(\delta^{-\alpha} y, T), \delta > 0\} \geqq \text{Min}_\delta \{\delta: (\delta^{1-\alpha} x') \in V(\delta^{-\alpha} y, T), \delta > 0\}, \alpha \in A.$$

3. Similarly, the function $D(x, y, T, \alpha)$ is increasing in y because, from assumptions A2 and A3, $y \geqq y'$ implies

$$\text{Min}_\delta \{\delta: (\delta^{1-\alpha} x) \in V(\delta^{-\alpha} y, T), \delta > 0\} \geqq \text{Min}_\delta \{\delta: (\delta^{1-\alpha} x) \in V(\delta^{-\alpha} y', T), \delta > 0\}, \alpha \in A.$$

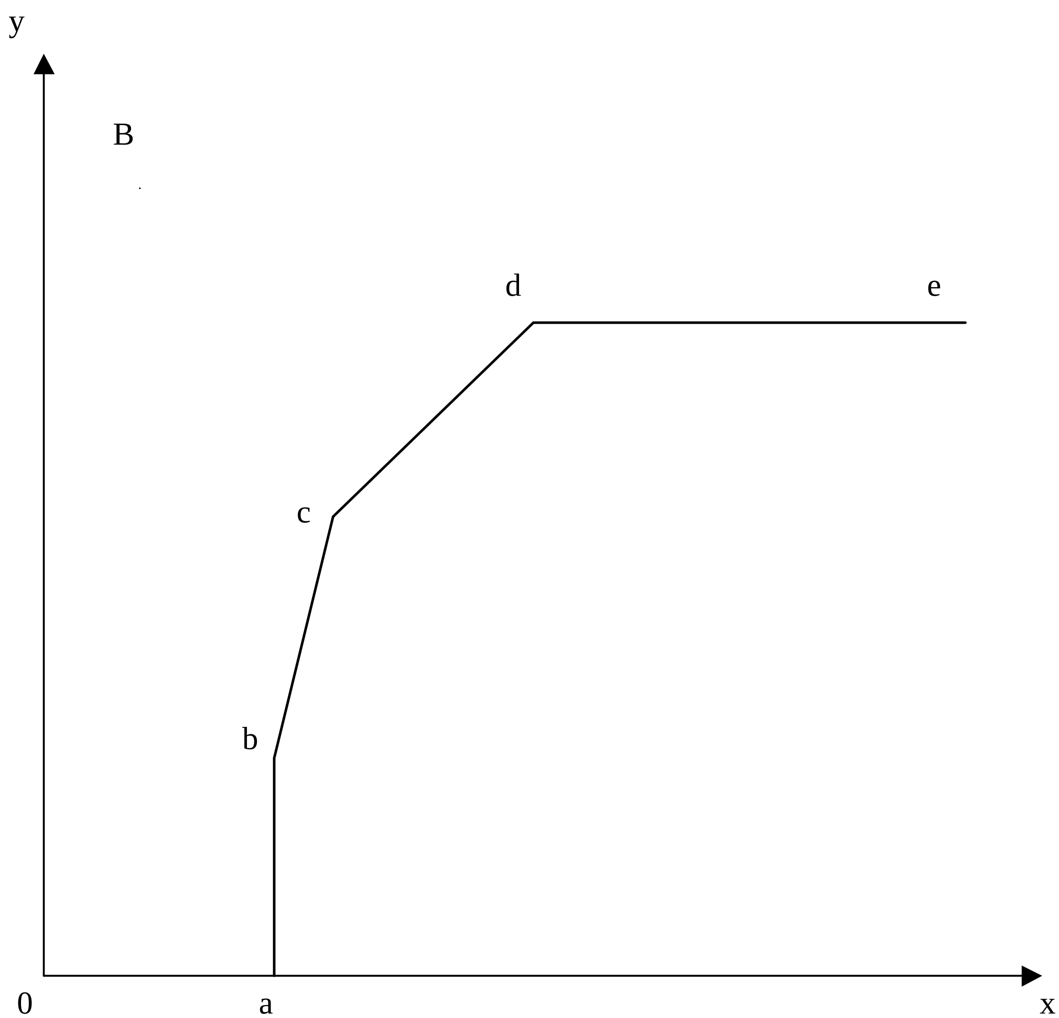
4. To show the lower semi-continuity of $D(x, y, T, \alpha)$ in (x, y), it suffices to show that the sets $\sigma(\alpha) = \{(x, y): D(x, y, T, \alpha) \le \beta\}$ are closed for all $\beta \in R$ (e.g., Shephard, p. 296). If $\beta \le 0$, $\sigma(\beta)$ is empty and thus closed. If $\beta > 0$, the almost homogeneity of $D(x, y, \cdot)$ implies that $\sigma(\beta) = \{(x, y): D(\beta^{1-\alpha} x, \beta^{-\alpha} y, T, \alpha) \leqq 1\}$. From proposition 1, it follows that $\sigma(\beta) = \{(x, y): (\beta^{1-\alpha} x) \in V(\beta^{-\alpha} y, T)\} = \{(x, y)\}: (\beta^{1-\alpha} x, \beta^{-\alpha} y) \in T\}$. Since T is a closed set by assumption A4, it follows that $\sigma(\beta)$ is closed for all $\beta > 0$.

5. Consider any two feasible values for $\alpha \in A$, α_0 and α_1 satisfying $\alpha_1 \geqq \alpha_0$. Under CRTS, choosing $\lambda = \delta^{\alpha_1-\alpha_0}$ and $\delta > 0$, it follows from equation (1a) that $V(\delta^{\alpha_1-\alpha_0} \delta^{-\alpha_1} y, T) = \delta^{\alpha_1-\alpha_0} V(\delta^{-\alpha_1} y, T)$. This implies $\{\delta: (\delta^{1-\alpha_0} x) \in V(\delta^{-\alpha_0} y, T), \delta > 0\} = \{\delta: (\delta^{1-\alpha_1} x) \in V(\delta^{-\alpha_1} y, T), \delta > 0\}$. From (2), this yields $D(x, y, T, \alpha_0) = D(x, y, T, \alpha_1)$.

Under IRTS, consider equation (1b) with $\lambda = \delta^{\alpha_1 - \alpha_0}$, $\alpha_1 \geqq \alpha_0$. Let $0 < \delta \leqq 1$. It follows from (1b) that $V(\delta^{\alpha_1-\alpha_0}\, \delta^{-\alpha_1}\, y, T) \subseteq \delta^{\alpha_1-\alpha_0}\, V(\delta^{-\alpha_1}\, y, T)$. This implies $\{\delta: (\delta^{1-\alpha_0}\, x) \in V(\delta^{-\alpha_0}\, y, T), \delta > 0\} \subseteq \{\delta: (\delta^{1-\alpha_1}\, x) \in V(\delta^{-\alpha_1}\, y, T), \delta > 0\}$ for $\delta \leqq 1$. From (2) and proposition 1, this yields $D(x, y, T, \alpha_0) \geqq D(x, y, T, \alpha_1)$ for $x \in V(y, T)$.

Finally, under DRTS, consider equation (1c) with $\lambda = \delta^{\alpha_1-\alpha_0}$, $\alpha_1 \geqq \alpha_0$. Let $0 < \delta \leqq 1$. It follows from (1c) that $V(\delta^{\alpha_1-\alpha_0}\, \delta^{-\alpha_1}\, y, T) \supseteq \delta^{\alpha_1-\alpha_0}\, V(\delta^{-\alpha_1}\, y, T)$. This implies $\{\delta: (\delta^{1-\alpha_0}\, x) \in V(\delta^{-\alpha_0}\, y, T), \delta > 0\} \supseteq \{\delta: (\delta^{1-\alpha_1}\, x) \in V(\delta^{-\alpha_1}\, y, T), \delta > 0\}$ for $\delta \leqq 1$. From (2) and proposition 1, this yields $D(x, y, T, \alpha_0) \leqq D(x, y, T, \alpha_1)$ for $x \in V(y, T)$.

Figure 1- A Representation of Production Function



Footnotes

1. Our approach can be seen as a special case of the Luenberger-Briec's approach, with the direction of inputs-outputs rescaling being controlled by the parameter α. In contrast with the more general Luenberger shortage function, our approach generates efficiency indexes that can be intuitively interpreted as ratios of average costs or average revenues (ray-average costs or ray-average revenues in a multi-input multi-output framework).
2. The following notation is used throughout the paper. R^n denotes the Euclidian space of dimension n. For $x = (x_1, ..., x_n) \in R^n$, $x \geqq 0$ means $x_i \geqq 0$, $i = 1, ..., n$. Alternatively, $x \geq 0$ means $x \geqq 0$ and $x \neq 0$. Finally, $x = (x_1, \ldots, x_n) \in R_+^n$ means $x \geqq 0$.
3. Note that these assumptions are rather weak. For example, they do not include the convexity of the set T. This suggests that the analysis presented below can be expected to hold under fairly general conditions.
4. For some alternative definition of return to scale, see Baumol et al. (p. 55) or Färe et al. (1988).
5. Note that a weaker attainability axiom has been introduced by Teusch (see Teusch; Färe and Mitchell).
6. Note that point B in Figure 1 is necessarily associated with technical change. Indeed, point B is above the line (abcde). If technically feasible, it must correspond to a "better" technology compared to the reference technology T represented by the production frontier (abcde). The issue raised by Figure 1 would therefore be relevant in the analysis of productivity and technical change (Caves et al., 1982b).
7. One possible extension of (2) would be to consider that only a subset of netputs are being rescaled. For example, if some inputs are considered fixed in the short run, then only variable inputs and outputs could be rescaled as in (2). This would generate "partial measures" conditional on the levels of fixed inputs. In the analysis presented below, this would appropriate in the measurement of technical and allocative efficiency. However, this would not make much sense in the analysis of scale efficiency, which is typically associated with long run equilibrium.
8. Briec defines the "Farrell equiproportional distance" as

   $$B = \text{Max}_s \{s: ((1-s)\, x) \in V((1+s)\, y, T)\}.$$

   This is a problem similar to (2) if $(1-s) = \delta^{1-\alpha}$ and $(1+s) = \delta^{-\alpha}$. It follows that the two optimization problems are equivalent when α is chosen to satisfy $2 = D(x, y, T, \alpha)^{-\alpha} + D(x, y, T, \alpha)^{1-\alpha}$. Then, the following relationships exist between the two measures: $B = D^{-\alpha} - 1 = 1 - D^{1-\alpha}$, and $D = (1-B)/(1+B)$.
9. Our approach is also related to the Luenberger-Briec's approach. To see this, consider the Luenberger

(1992b, 1995) shortage function

$$\sigma(x, y, T, g_x, g_y) = \mathrm{Inf}_s\{s: (x + s\, g_x) \in V(y - s\, g_y)\},$$

where $g_x \in R_+^n$ is a (n×1) vector, $g_y \in R_+^m$ is a (m×1) vector, and $(g_x', g_y') \neq 0$. The shortage function σ measures how far "short" is (x, y) from being on the frontier technology in the direction g_x on the input side and g_y on the output side. This differs from (2) in two ways. First, in general, the directions (g_x', g_y') in the Luenberger-Briec approach are chosen exogenously. In contrast, the rescaling in (2) is always toward zero. We will see below that this rescaling toward zero (a feature not shared with the Luenberger-Briec approach) will help generate efficiency indexes that have simple and intuitive economic interpretations. Second, the Lunberger-Briec moves toward the frontier technology can take place in <u>any</u> direction $(g_x', g_y') \geq 0$. As such, it is more general than the rescaling considered in (2). However, particular choices for (g_x', g_y') can show how the two measures are related. To see this, consider the specific directions $g_x = [(1 - b)\, x]$ and $g_y = (b\, y)$, where b is a scalar, $0 \leq b \leq 1$. Then, for each α in (2), there exists a scalar b such that the moves from (x, y) are in the same directions as in (2). In this situation, comparing the shortage function with (2) gives: $1 - \sigma\, b = D^{-\alpha}$, and $1 + \sigma\,(1\text{-}b) = D^{1-\alpha}$. This implies the following relationship between our generalized distance function D in (2) and Luenberger's shortage function σ: $D = (1 - \sigma\, b)^{-1/\alpha} = (1 + \sigma\,(1\text{-}b))^{-1/(\alpha-1)} = (1 + \sigma\,(1\text{-}b))/(1 - \sigma\, b)$, and $\sigma = D^{1-\alpha} - D^{-\alpha}$. When $\alpha = b = 0$ (where only inputs are rescaled), this yields $\sigma(x, y, T, x, 0) = D(x, y, T, 0) - 1$. Alternatively, when $\alpha = b = 1$ (where only outputs are rescaled), this gives $\sigma(x, y, T, 0, y) = 1 - D(x, y, T, 1)$.

10. A function F(x, y) is almost homogeneous of degrees r, s and t in x and y if and only if $F(\lambda^r x, \lambda^s y) = \lambda^t F(x, y)$ for any scalar $\lambda > 0$ (see Aczél, p. 231; Lau).

11. Under VRTS where the technology exhibits DRTS <u>and</u> IRTS (in different regions), the solution for α in (3) would depend on (x, y). In this case, choosing α for different values of (x, y) would require modifying (3). For example, given a sample of N observations on inputs and outputs (x^i, y^i) associated with a given technology T, i = 1, …, N, as an alternative to (3), one may choose α as follows

$$\alpha \in A^* = \mathrm{argmin}_\alpha \{ \Sigma_{i=1}^N |D(x^i, y^i, T, \alpha) - 1|: \alpha \in A\}.$$

12. We assume that the cost minimization problem has a solution.

13. To simplify the notation through the rest of the paper, we will use "TE" to mean TE(x, y, T, α).

14. We assume that the revenue maximization problem has a solution.

15. We assume that the profit maximization problem has a solution. Note that this rules out any technology exhibiting (global) constant or increasing returns to scale.

16. Note that a distinction is sometimes made in the literature between scale efficiency and size efficiency. We do not elaborate on this distinction here. Instead, we emphasize alternative measures of the firm efficiency of the scale of operation using cost, revenue, and profit.

17. Note that the ray-average cost function can be interpreted as a multi-product generalization of the single-product "average cost function" commonly found in microeconomic textbooks (e.g., Chambers). See Baumol et al. for a good discussion of the properties of the ray-average cost function.

18. We implicitly assume that that the infimum in (13) exists. It does if there is a scalar $k > 0$ such that $(r'x/\lambda) \geq k$ for all $x \geq 0$, $y \geq 0$, $x \in V(\lambda y, T)$, and $\lambda > 0$. Note that such a condition imposes rather mild restrictions on technology since it allows for increasing, constant, or decreasing return to scale at any point of the feasible region. This condition simply eliminates the possibility of having a zero "average cost".

19. The relationship between measuring return to scale from the production technology versus the cost function has been discussed in detail in the literature (e.g., see Baumol et al.).

20. Note that, in a single-product context, ray-average revenue is closely related to the ray-average product commonly found in microeconomic textbooks (e.g., Chambers, p. 25). Ray-average revenue can (loosely) be interpreted as a (multi-product) extension of such a concept applied to revenue.

21. We implicitly assume that that a finite supremum in (16) exists. It does if there is a scalar $k < \infty$ such that $(p'y/\lambda) \leq k$ for all $x \geq 0$, $y \geq 0$, $(\lambda x) \in V(y, T)$, and $\lambda > 0$. Note that such a condition imposes rather mild restrictions on technology since it allows for increasing, constant, or decreasing return to scale at any point of the feasible region. This condition simply eliminates the possibility of having an infinite "average revenue".

22. Note that the analysis below could alternatively be presented using the revenue-to-cost ratio $p'y/(r'x)$, generating the revenue-cost function $RC(r, p, T) = \mathrm{Sup}_{x,y} \{p'y/(r'x): x \in V(y, T), x \geq 0, y \geq 0\}$, with $RC(r, p, T) = 1/CR(r, p, T)$.

23. We implicitly assume that a positive infimum in (18) exists. It does if there is a scalar $k > 0$ such that $r'x/(p'y) \geqq k$ for all $x \geq 0$, $y \geq 0$, and $x \in V(y, T)$. This is a rather mild condition since it allows for increasing, constant, or decreasing return to scale at any point of the feasible region. It simply eliminates the possibility of having a zero "average cost" or an infinite "average revenue".

24. Note that, under differentiability of D(x, y, T), the ratio $C^*(r, p, T)/R^*(r, p, T)$ is the "scale elasticity" S for a profit maximizing firm, where $S = -[\sum_{i=1}^{n} x_i\, \partial D/\partial x_i]/[\sum_{j=1}^{m} y_j\, \partial D/\partial y_j]$, evaluated at $x^*(r, p, T)$ and $y^*(r, p, T)$ (e.g., Baumol et al., p. 55-56). The scale elasticity S measures the proportional change in outputs y on the production frontier due to a (small) proportional change in inputs x. It has been commonly used in empirical work. This establishes a useful linkage between this measure and our analysis.