



Centre for Efficiency and Productivity Analysis

**Working Paper Series
No. WP06/2011**

A Scale Elasticity Measure for Directional Distance Function and its Dual

Valentin Zelenyuk

Date: August 2011

**School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia**

ISSN No. 1932 - 4398

A Scale Elasticity Measure for Directional Distance Function and its Dual

Valentin Zelenyuk*

Abstract

In this paper we introduce a scale elasticity measure based on directional distance function for multi-output-multi-input technologies and explore its fundamental properties. Specifically, we derive necessary and sufficient condition for equivalence of the scale elasticity measure based on the directional distance function with the input oriented and output oriented scale elasticity measures. We also establish duality relationship between the scale elasticity measure based on the directional distance function with a scale elasticity measure based on the profit function. This result is theoretical, yet it is also valuable for empirical researchers as it provides a testable analytical condition for when (and only when) the alternative primal and dual definitions of scale elasticity for multi-output-multi-input technologies yield equivalent conclusions about economies or diseconomies of scale.

Key words: Scale elasticity, production theory, distance functions.

JEL Code: D20, D24.

This version: August 5, 2011

Acknowledgement: I would like to thank Rolf Färe, Jeff Kline, Antonio Peyrache, Chris O'Donnell, Prasada Rao and Natalya Zelenyuk for fruitful discussions and valuable feedback to this paper. I remain to be solely responsible for the views expressed in this article.

* School of Economics and Centre for Efficiency and Productivity Analysis, The University of Queensland, Australia. *Address:* 530, Colin Clark Building (39), St Lucia, Brisbane Qld 4072, AUSTRALIA; *e-mail:* v.zelenyuk@uq.edu.au; *tel:* + 61 7 3346 7054;

1. Introduction

Since its inception by Chambers, Chung and Färe (1996, 1998), and earlier inspiration from the fundamental works by Luenberger (1992, 1994, 1995), the directional distance function or, Luenberger's benefit function, has been gaining increasing popularity both in theoretical and empirical studies. While being a very convenient tool for characterizing, estimating and analyzing multi-output-multi-input technologies as well as for measuring welfare effects, the directional distance function and the benefit function found many theoretical and empirical uses including theoretical international trade theory, productivity, efficiency and economic growth analysis, environmental shadow price estimation, etc.ⁱ One of the most important aspects in applied analysis of firms is measurement of economies and diseconomies of scale and here, in this article, we focus on this aspect within the framework of directional distance function.

The aspect of measuring economies or diseconomies of scale was explored in many studies, including the seminal works of Hanoch (1975), Panzar and Willig (1977), Färe, Grosskopf and Lovell (1986), to mention just a few, and we extend these ideas in the present work to the context of directional distance function (DDF). In particular, within the standard production economics theory framework (following Shephard (1953, 1970) and Färe and Primont (1995)), we first introduce a scale elasticity measure based on the DDF and then derive the necessary and sufficient condition for the equivalence between our measure and the existing scale elasticity measures based on the input oriented and output oriented Shephard's distance functions. We also establish duality relationship between the scale elasticity measure based on the directional distance function and a scale elasticity measure based on the profit function.

Since the choice of characterization of technology in practice is often made arbitrary, an empirical value of our theoretical result is that it provides researchers with an analytical condition (which is necessary and sufficient) that could be verified empirically with available data and an

appropriate estimator. In practice, testing this condition can help clarifying if the researcher's results about scale elasticity estimates would be different if one were to use a different characterization of technology for this same data set.

The rest of this paper is structured as following: Section 2 outlines the various approaches for technology characterization involved in this work. Section 3 outlines the alternative definitions of scale elasticity and, in particular, introduces our scale elasticity measure based on the DDF. Section 4 states, proves and briefly discusses the primal results. Section 5 states, proves and briefly discusses the dual results. Section 6 concludes.

2. Characterizations of Technology

To facilitate our formal discussion, let $x = (x_1, \dots, x_N)' \in \mathfrak{R}_+^N$ be a vector of inputs and $y = (y_1, \dots, y_M)' \in \mathfrak{R}_+^M$ be a vector of outputs, and assume that the production technology of a firm is characterized by the technology set $T \subset \mathfrak{R}_+^N \times \mathfrak{R}_+^M$, defined as

$$T \equiv \{(x, y) \in \mathfrak{R}_+^N \times \mathfrak{R}_+^M : y \in \mathfrak{R}_+^M \text{ is producible from } x \in \mathfrak{R}_+^N\}. \quad (1)$$

We will assume that technology satisfies ‘standard regularity conditions’ of production theory such as

- (i) “no free lunch” $((0, y) \notin T \text{ for any } y \neq 0)$,
- (ii) “doing nothing is possible” $((x, 0) \in T \text{ for any } x \in \mathfrak{R}_+^N)$,
- (iii) the set T is a closed set,
- (iv) the output sets of T (defined by $P(x) := \{y : (x, y) \in T\}$, $x \in \mathfrak{R}_+^N$) are bounded for any $x \in \mathfrak{R}_+^N$,

(v) technology set T satisfies ‘free disposability’ for all inputs and all outputs, i.e.,

$$(x, y) \in T \Rightarrow (x', y') \in T, \forall y' \leq y \text{ and } \forall x \leq x'.$$

For details of these regularity conditions and resulting properties see Färe and Primont (1995) as well as Chambers et al., (1996, 1998).

In a single output case, a common approach to completely characterize the technology set T is to use the production function $f : \mathfrak{R}_+^N \rightarrow \mathfrak{R}_+^1$ defined as

$$f(x) \equiv \max\{y : (x, y) \in T\}, \quad (2)$$

To characterize technology in a multi-output-multi-input context, one can use many appropriate functions, most popular of which appear to be the Shephard’s distance functions, which we will involve later in this paper as well. For this, recall that the output oriented Shephard’s (1970) distance function $D_o : \mathfrak{R}_+^N \times \mathfrak{R}_+^M \rightarrow \mathfrak{R}_+^1 \cup \{+\infty\}$ is defined as

$$D_o(x, y) \equiv \inf\{\theta > 0 : (x, y/\theta) \in T\}, \quad (3)$$

while the Shephard’s (1953) input distance function $D_i : \mathfrak{R}_+^M \times \mathfrak{R}_+^N \rightarrow \mathfrak{R}_+^1 \cup \{+\infty\}$ is defined as

$$D_i(y, x) \equiv \sup\{\theta > 0 : (x/\theta, y) \in T\}. \quad (4)$$

It is well known that under fairly mild conditions on technology, both functions possess many useful properties, in particular, they both completely characterize technology set T in the sense that

$$D_o(x, y) \leq 1 \Leftrightarrow D_i(y, x) \geq 1 \Leftrightarrow (x, y) \in T \quad (5)$$

Our focus in this work is on the directional distance function that is more general than the Shephard's distance functions, and includes them as special cases. Recall that the directional distance function $D_d : \mathfrak{R}_+^N \times \mathfrak{R}_+^M \rightarrow \mathfrak{R}_+^1$ is defined as

$$D_d(x, y | d_x, d_y) \equiv \sup \{ \beta : (x, y) + \beta(-d_x, d_y) \in T \}, \quad (6)$$

where $d = (d_x, d_y) \in \mathfrak{R}_+^N \times \mathfrak{R}_+^M$ is a non-zero direction vector specified by the researcher.

Many useful properties of this function were established in Chambers, Chung and Färe (1996, 1998). The most important property for us is that, under the standard regularity condition (i)-(v), the directional distance function completely characterizes the technology set T , in the sense that

$$D_d(x, y | d_x, d_y) \geq 0 \Leftrightarrow (x, y) \in T. \quad (7)$$

An important advantage of this characterization of technology over others is that it is dual to the profit function,ⁱⁱ formally defined as

$$\pi(w, p) \equiv \max_{x, y} \{ p'y - w'x : (x, y) \in T \}. \quad (8)$$

In the next section we will consider various measures of scale elasticity based on the different primal characterizations of technology.

3. Primal Measures of Scale Elasticity

First, recall that in the case of single output technology, a commonly used measure of scale elasticity cited in many economics textbooks is given by

$$e(x) \equiv \left. \frac{\partial \log f(\lambda x)}{\partial \log \lambda} \right|_{\lambda=1} = \sum_{i=1}^N \frac{\partial f(x)}{\partial x_i} \frac{x_i}{f(x)} = \nabla_x' f(x)x / f(x). \quad (9)$$

where $f(x)$ is the production function defined in (2) that is assumed to be differentiable (at least once) and where $\nabla_x' f(x) \equiv (\partial f(x) / \partial x_1, \dots, \partial f(x) / \partial x_N)$ is its gradient.

In their seminal works, Färe, Grosskopf and Lovell (1986) and Färe and Primont (1995), modifying ideas of Hanoch (1975), Panzar and Willig (1977), generalized the scale elasticity measure in (9) to the multiple output case by employing the Shephard's distance functions. In particular, their *output oriented measure of scale elasticity* can be defined as

$$e_o(x, y) = \left. \frac{\partial \log \theta}{\partial \log \lambda} \right|_{\theta=1, \lambda=1}, \quad \text{such that} \quad D_o(\lambda x, \theta y) = 1. \quad (10)$$

Alternatively, one could measure returns to scale based on the input oriented distance function (4), defining the *input oriented measure of scale elasticity* as

$$e_i(y, x) = \left. \frac{\partial \log \lambda}{\partial \log \theta} \right|_{\theta=1, \lambda=1}, \quad \text{such that} \quad D_i(\theta y, \lambda x) = 1. \quad (11)$$

Intuitively, both measures are trying to gauge the scale elasticity by looking at the relationship between *equi-proportional* changes in *all* inputs with *equi-proportional* changes in *all* outputs, but they do it by using different characterizations of technology and, in some sense, in ‘orthogonal’ directions. In the definitions above, notice that while the output oriented measure of scale elasticity in (10) is exactly the same as that defined in Färe, Grosskopf and Lovell (1986) and in Färe and Primont (1995), our input oriented measure of scale elasticity in (11) is the reciprocal of their input oriented measure of scale elasticity. The issue which one to use is a matter of taste or convenience. Here, we chose to use (11) for convenience, to preserve the analogy that exists between the output oriented and input oriented distance functions. Indeed, note that because, the output oriented distance function is measuring (maximal equi-proportional) increase in outputs feasible at a given level of inputs, it is natural to define the output oriented scale elasticity measure as a ratio of equi-proportional percent change in outputs given equi-proportional

percent increase in all inputs, and this is exactly how measure in (10) is stated. Analogously, because the input oriented distance function is measuring (maximal equi-proportional) decrease in inputs that makes a given level of outputs feasible, it is natural to define the input oriented scale elasticity measure as a ratio of equi-proportional percent change in inputs to equi-proportional percent increase in all outputs, and this is exactly how measure in (11) is formulated. Notice, however, that while (10) has conventional interpretation (e.g., values bigger than 1 indicate about increasing returns to scale at the point of measurement), measure in (11) has ‘reciprocal’ interpretation. Indeed, for (11) to indicate increasing returns to scale it must yield a value below 1—because increasing returns to scale implies that increase of outputs by some (infinitesimal) percentage requires increase of inputs by an even *smaller* percentage. Of course, one could always convert (11) to the same units of measurement as (10) by taking its reciprocal and, perhaps, this was one of the motivations for Färe, Grosskopf and Lovell (1986) and Färe and Primont (1995) when they defined their version.

Let us now turn to the scale measurement based on the DDF. While there might be different ways of defining a scale elasticity measure based on the DDF, it seems natural to define it on analogy with the measures defined in (10) and (11) and we will do so here. That is, the measure of scale elasticity based on the DDF, for a given direction (d_x, d_y) , is defined as

$$e_d(x, y | d_x, d_y) = \frac{\partial \log \theta}{\partial \log \lambda} \bigg|_{\theta=1, \lambda=1}, \quad \text{such that} \quad D_d(\lambda x, \theta y | d_x, d_y) = 0. \quad (12)$$

Intuitively, and analogously to measures in (10) and (11), the scale elasticity measure in (12) is telling us about *equi-proportional* percent change in *all* outputs due to *equi-proportional* change in *all* inputs.ⁱⁱⁱ

Now, several natural questions arise: ‘What is a relationship between the scale elasticity measure based on the DDF and the scale elasticity measures based on the input and output oriented

Shephard's distance functions?" In particular, are these measures equivalent? Always? Under what conditions? Since in general, the Shephard's distance functions and the DDF do *not* have explicit one-to-one equality relationship with each other (except for peculiar cases such as CRS, or very special directions as $(0, \beta y)$ or $(\alpha x, 0)$ for some non-zero scalars α and β) the answer is not straightforward. In the next section we establish one important result—we state and prove a new theorem about such relationship, outlining the necessary and sufficient condition for this relationship. As many proofs in economics, the derivation in the proof of our results is facilitated with the Lagrangian function and the envelop theorem.

4. Primal Equivalences

To establish the equivalence of interest, we will focus on the subset of the technology set where all distance functions we considered suggest that the input-output allocation is technically efficient, i.e., we focus on points belonging to $T^\partial \subset T$, where

$$T^\partial = \{(x, y) \in T : D_o(x, y) = 1, D_i(y, x) = 1, D_d(x, y | d_x, d_y) = 0\}.$$

If there is an interest in an input-output allocation in the interior of T , then one must choose the direction the inefficiency is to be measured along, then find the projection of this input-output allocation onto the frontier for this direction and then measure the scale properties at that projection. Importantly, note that for the technically inefficient points (i.e., where $(x, y) \notin T^\partial$), even if one uses the same (input or output or directional) distance function, the values of scale elasticity might differ substantially depending on the direction chosen for efficiency measurement, and can even suggest opposite conclusions. Therefore, in general, for input-output allocations where at least one of the distance functions suggests that $(x, y) \notin T^\partial$, one

cannot guarantee the equivalence except for the very special cases, and so in the derivation of our results we will focus only on $(x, y) \in T^\partial$.

Theorem 1.

Given definitions (1), (3), (4), (6), (10), (11) and (12), standard regularity conditions of production theory (i)-(v) and assuming that at a point of interest $(x, y) \in T^\partial$ and where the functions $D_d(x, y | d_x, d_y)$, $D_i(y, x)$ and $D_o(x, y)$ are differentiable at least once w.r.t. each element of (x, y) , we have:

$$e_d(x, y | d_x, d_y) = e_o(x, y) = 1/e_i(y, x) \quad (13)$$

if and only if

$$\nabla'_y D_i(y, x)y \neq 0 \quad \text{and} \quad \nabla'_y D_d(x, y | d_x, d_y)y \neq 0 \quad \text{and} \quad \nabla'_x D_d(x, y | d_x, d_y)x \neq 0. \quad (14)$$

Proof To prove the necessity of (14), assume that (13) is true and then note that the scale elasticity in (12), can be rewritten as

$$e_d(x, y | d_x, d_y) = -\frac{\nabla'_x D_d(x, y | d_x, d_y)x}{\nabla'_y D_d(x, y | d_x, d_y)y} \quad (15)$$

This result follows from application of the implicit function theorem, applied to (12), conditioning on the chosen direction (d_x, d_y) and, clearly requires condition $\nabla'_y D_d(x, y | d_x, d_y)y \neq 0$ to be valid. Similarly, note that also by using the implicit function theorem, we can rewrite definitions in (10) and (11), respectively, also in a more compact form

$$e_o(x, y) = -\nabla'_x D_o(x, y)x, \quad (16)$$

and

$$e_i(y, x) = -\nabla_y' D_i(y, x)y. \quad (17)^{iv}$$

And so, if (13) is true then it also must be true that $-\nabla_y' D_i(y, x)y \neq 0$. This, in turn, implies that $\nabla_x' D_d(x, y | d_x, d_y)x = 0$ is ruled out because this could happen only when $e_d(x, y | d_x, d_y) = e_i(x, y) = 1/e_i(y, x) = 0$, but the last equality of it is ruled out because of (17) together with the fact that $-\nabla_y' D_i(y, x)y \neq 0$ must be true. This concludes the proof of necessity of (14) for (13).

To prove the *sufficiency* part, assume (14) is true and note that, due to (3) and (5), we can rewrite the output distance function as following:

$$D_o(x, y) = \inf\{\theta > 0 : D_d(x, y/\theta | d_x, d_y) \geq 0\}. \quad (18)$$

The corresponding Lagrangian function for this optimization problem can then be written as

$$L(\theta, \gamma | x, y) = \theta - \gamma(D_d(x, y/\theta | d_x, d_y) - 0), \quad (19)$$

and the associated f.o.c. is given by

$$\nabla_\theta L \Big|_{\substack{\theta=\theta^* \\ \gamma=\gamma^*}} = 1 - \gamma^* \nabla_{y/\theta}' D_d(x, y/\theta^* | d_x, d_y)y(-1/(\theta^*)^2) = 0, \quad (20)$$

and

$$\nabla_\gamma L \Big|_{\substack{\theta=\theta^* \\ \gamma=\gamma^*}} = D_d(x, y/\theta^* | d_x, d_y) = 0, \quad (21)$$

where $\theta^* = \theta(x, y)$ and $\gamma^* = \gamma(x, y)$ are the optimal solutions to optimization problem (18).

Now, because θ^* is a solution to (18), its value must be equal to unity, and therefore (20)

reduces to

$$\gamma^* \nabla_y' D_d(x, y | d_x, d_y) y = -1. \quad (22)$$

On the other hand, note that the envelope theorem applied to (19), tells us that

$$\nabla_x' D_o(x, y) = \nabla_x' L(\theta^*, \gamma^* | x, y) = -\gamma^* \nabla_x' D_d(x, y | \theta^* | d_x, d_y). \quad (23)$$

Now, post-multiplying both sides of (23) by the vector of inputs and by (-1), and using again our knowledge that at the optimum we must have $\theta^* = 1$, we can rewrite (23) as

$$-\nabla_x' D_o(x, y) x = \gamma^* \nabla_x' D_d(x, y | d_x, d_y) x. \quad (24)$$

And so, noting that the l.h.s. of (24) is exactly the output oriented scale elasticity, according to (16), and combining this result with (22), and with our assumption that $\nabla_y' D_d(x, y | d_x, d_y) y \neq 0$, we obtain

$$e_o(x, y) = -\frac{\nabla_x' D_d(x, y | d_x, d_y) x}{\nabla_y' D_d(x, y | d_x, d_y) y} = e_d(x, y | d_x, d_y). \quad (25)$$

Along the same logic as above, we can also rewrite the input distance function as

$$D_i(y, x) = \inf\{\lambda > 0 : D_d(x/\lambda, y | d_x, d_y) \geq 0\}, \quad (26)$$

and so the corresponding Lagrangian function for this optimization problem can be written as

$$L(\theta, \lambda | x, y) = \lambda - \delta(D_d(x/\lambda, y | d_x, d_y) - 0), \quad (27)$$

and the associated f.o.c. is given by

$$\nabla_\lambda L \Big|_{\substack{\lambda=\lambda^* \\ \delta=\delta^*}} = 1 - \delta^* \nabla_{x/\lambda} D_d(x/\lambda^*, y | d_x, d_y) x (-1/(\lambda^*)^2) = 0, \quad (28)$$

and

$$\nabla_{\delta} L \Big|_{\substack{\lambda=\lambda^* \\ \delta=\delta^*}} = D_d(x/\lambda^*, y | d_x, d_y) = 0, \quad (29)$$

where $\lambda^* = \lambda(x, y)$ and $\delta^* = \delta(x, y)$ are the optimal solutions to (26). Similarly as above, because λ^* is a solution to (26), its value must be equal to unity, and therefore (28) reduces to

$$\delta^* \nabla'_x D_d(x, y | d_x, d_y) x = -1. \quad (31)$$

On the other hand, note that the envelope theorem applied to (27), tells us that

$$\nabla'_y D_i(y, x) = \nabla'_y L(\lambda^*, \delta^* | x, y) = -\delta^* \nabla'_y D_d(x, y/\lambda^* | d_x, d_y). \quad (32)$$

Now, post-multiplying both sides of (32) by the vector of inputs and by (-1), and using again our knowledge that $\lambda^* = 1$, we can rewrite (32) as

$$-\nabla'_y D_i(y, x) x = \delta^* \nabla'_y D_d(x, y | d_x, d_y) y. \quad (33)$$

And so, noting that the l.h.s. of (33) is exactly the input oriented scale elasticity, and combining this result with (31), and with assumption that $\nabla'_x D_d(x, y | d_x, d_y) x \neq 0$, we obtain

$$e_i(y, x) = -\nabla'_y D_i(y, x) x = -\frac{\nabla'_y D_d(x, y | d_x, d_y) y}{\nabla'_x D_d(x, y | d_x, d_y) x} = (e_d(x, y | d_x, d_y))^{-1}. \quad (34)$$

Q.E.D.

It is worth noting here that for strictly positive input output allocations, i.e., $(x, y) \in \mathfrak{R}_{++}^{N+M}$, the necessary and sufficient condition (14) is simplified to

$$\nabla'_y D_i(y, x) \neq 0 \quad \text{and} \quad \nabla'_y D_d(x, y | d_x, d_y) \neq 0 \quad \text{and} \quad \nabla'_x D_d(x, y | d_x, d_y) \neq 0. \quad (35)$$

Intuitively, as one might expect, the theorem we stated and proved above tells us that, under fairly mild conditions, the three scale elasticity formulas we stated above measure the same property of technology equivalently. Intuitively, the necessary and sufficient condition (14) states that, at the particular points where scale elasticity is to be measured, the gradients of the input and output distance functions and the gradients of the directional distance function with respect to output and with respect to inputs shall *not* be orthogonal to the output and input vectors, respectively. Moreover, as stated in (35), in the special case of measuring at strictly positive input-output combination, the necessary and sufficient condition reduces to requirement that at least one partial derivative of the input distance function and at least one partial derivative of the output distance function is positive.

On one hand, in technical terms, the requirement (14) simply ensures not running into situation of division by zero and so by this condition, we ensure that at a given point of measurement, none of the two measures of scale elasticity explodes and none degenerates to zero, and then (and only then) they give equivalent information about the scale of technology at that particular point. On the other hand, the necessary and sufficient condition (14) also has an economic meaning: it says that an increase in *all* inputs (outputs) by the same *proportion* necessitates some proportional, non-zero finite change in *all* outputs (inputs).

Importantly, note that the theorem above outlines the necessary and sufficient restriction on technology that is a local (rather than a global) requirement, i.e., it is about particular input-output allocations at which elasticity is to be measured. In other words, while at some points the equivalence may happen to fail to be true, it still might hold for most of the points of interest in practice (e.g., at the average, median, certain quantiles of interest, etc.), and so, in practice, it might be enough to verify condition (14) at these points of interest only. In this respect, out

theoretical result attains an empirical importance in modelling multi-output-multi-input technologies, where it became very popular to estimate various distance functions.^v Note that in empirical analysis some researchers choose output-oriented Shephard's distance function, while others choose the input-oriented Shephard's distance function and yet others give preference to the directional distance function. Such choices are often arbitrary and it is not always clear whether results based on these alternative characterizations of technology would or should be the same or similar (due to some estimation noise), at least qualitatively. As a matter of fact, for some measurements it is well known that results would not be equivalent in general. For example, efficiency measurement would give equivalent results only under the case of a constant return to scale technology, which is a trivial case for our context. An empirical value of this paper is that it provides a testable theoretical condition on when such alternative approaches to modelling the production process should yield equivalent results, for the particular case of measuring scale elasticity, and this condition can be tested.

In the next section we discover another important equivalence result—equivalence between the DDF-based scale elasticity measure and a profit-based measure of scale elasticity, and so establishing a duality relationship between these two alternative measures.

5. Dual Equivalences

Various duality results for DDF have been established by Luenberger (1992), Chambers et al. (1996, 1998) and Färe and Primont (2006). Some duality implications for scale elasticity measurement were established by Färe, Grosskopf and Lovell (1986) and reinstated in Färe and Primont (1995), who show duality relationship of (10) and (11) to scale elasticity measures based on the revenue and cost functions, respectively. To our knowledge, duality relationship for the

scale elasticity based on the profit function with that based on the DDF has not been derived yet and this is the goal of this section.^{vi}

A measure of scale elasticity based on the profit function can be defined analogously to definitions in (10), (11) and (12), i.e., as

$$e_{\pi}(p, w) = \frac{\partial \log(\theta)}{\partial \log(\lambda)} \quad , \quad \text{such that} \quad \pi(\theta p, \lambda w) = \pi^{\circ} \quad , \quad (36)$$

where π° is some real constant (which can be zero, to satisfy the zero-profit condition). The intuition of this measure is similar to those we have for (10), (11) and (12), but with the dual meaning. Specifically, scale elasticity measure in (36) is, intuitively, telling us at which percentage rate should all the output prices change (equi-proportionately) given one per cent (equi-proportionate) change in all the input prices, such that the profit of the profit maximizing agent (e.g., firm) stays the same. The scale elasticity measure in (36) is particularly useful when researcher is operating with the profit function to characterize and analyse technology under assumption of optimal (profit-maximizing) behaviour of the analyzed firm. This framework is consistent with economic theory of firms as well as might be the only feasible approach when primal data (inputs and outputs) of a firm of interest is unavailable but researcher has dual (prices) data, as required by $\pi(p, w)$. In the next theorem, we establish relationship between $e_{\pi}(p, w)$ and $e_d(x, y | d_x, d_y)$.

Theorem 2.

Given definitions (1), (6), (8), (12) and (36), standard regularity conditions of production theory (i)-(v) and assuming that $D_d(x, y | d_x, d_y)$ and $\pi(w, p)$ are differentiable at least once w.r.t. each of their element, we have:

$$e_{\pi}(p, w) = e_d(x^*, y^* | d_x, d_y), \quad (37)$$

if and only if

$$\nabla'_x D_d(x^*, y^* | d_x, d_y) x^* \neq 0 \quad \text{and} \quad w' \nabla_w \pi(p, w) \neq 0, \quad (38)$$

and where

$$(x^*, y^*) \equiv (x(p, w), y(p, w)) \equiv \arg \max_{x, y} \{ p' y - w' x : (x, y) \in T \}. \quad (39)$$

Proof: To prove necessity of (38), assume (37) is true and this would require that

$\nabla'_x D_d(x^*, y^* | d_x, d_y) x^* \neq 0$. Moreover, using implicit function theorem, we can rewrite (36) as

$$e_{\pi}(p, w) = - \frac{p' \nabla_p \pi(p, w)}{w' \nabla_w \pi(p, w)} \quad (40)$$

which immediately requires that $w' \nabla_w \pi(p, w) \neq 0$, completing the proof of necessity of (38) for (37).

To prove sufficiency of (38) for (37), assume (38) is true and note that, in general, due to (6) and (7), we can rewrite the profit function as:

$$\pi(w, p) \equiv \max_{x, y} \{ p' y - w' x : D_d(x, y | d_x, d_y) \geq 0 \}, \quad (41)$$

The corresponding Lagrangian function for this optimization problem can then be written as

$$L(x, y, \rho | p, w, d_x, d_y) = p' y - w' x - \rho (D_d(x, y | d_x, d_y) - 0), \quad (42)$$

and so the system of equations defined by the associated first order condition is given by

$$\nabla_y L \Big|_{\substack{y=y^* \\ x=x^* \\ \rho=\rho^*}} = p' - \rho^* \nabla'_y D_d(x^*, y^* | d_x, d_y) = 0, \quad (43)$$

and

$$\nabla_x L \Big|_{\substack{y=y^* \\ x=x^* \\ \rho=\rho^*}} = w' - \rho^* \nabla_x' D_d(x^*, y^* | d_x, d_y) = 0, \quad (44)$$

and

$$\nabla_y L \Big|_{\substack{y=y^* \\ x=x^* \\ \rho=\rho^*}} = D_d(x^*, y^* | d_x, d_y) = 0, \quad (45)$$

where $y^* = y(p, w)$, $x^* = x(p, w)$, and $\rho^* = \rho(p, w)$ are the solutions to (41). Furthermore, rearranging (43) and (44), we get

$$p' = \rho^* \nabla_y' D_d(x^*, y^* | d_x, d_y) \quad (46)$$

and

$$w' = \rho^* \nabla_x' D_d(x^*, y^* | d_x, d_y) \quad (47)$$

which in turn imply that

$$p' y^* = \rho^* \nabla_y' D_d(x^*, y^* | d_x, d_y) y^* \quad (48)$$

and

$$w' x^* = \rho^* \nabla_x' D_d(x, y | d_x, d_y) x^* \quad (49)$$

Moreover, from the envelope theorem applied to (42), we get

$$\nabla_p \pi(p, w) = \nabla_p L(x^*, y^*, \rho^* | p, w, d_x, d_y) = y^*. \quad (50)$$

and

$$\nabla_w \pi(p, w) = \nabla_w L(x^*, y^*, \rho^* | p, w, d_x, d_y) = -x^*. \quad (51)$$

which are the Hotelling/Shephard's lemmas, and they in turn imply that

$$p' \nabla_p \pi(p, w) = p' y^* \quad \text{and} \quad w' \nabla_w \pi(p, w) = -w' x^*. \quad (52)$$

Therefore, assuming $w' \nabla_w \pi(p, w) \neq 0$ in (38) in its turn implies that $w' x^* \neq 0$, and so we can combine equations (48) and (49) to write

$$e_\pi(p, w) = -\frac{p' \nabla_p \pi(p, w)}{w' \nabla_w \pi(p, w)} = \frac{p' y^*}{w' x^*}. \quad (53)$$

On the other hand, combining (48) and (49), with assumption that $\nabla'_x D_d(x, y | d_x, d_y)x^* \neq 0$, we get

$$\frac{p'y^*}{w'x^*} = \frac{\nabla'_y D_d(x^*, y^* | d_x, d_y)y^*}{\nabla'_x D_d(x, y | d_x, d_y)x^*} = e_d(x^*, y^* | d_x, d_y). \quad (54)$$

Finally, combining (53) with (54) we arrive to (37), completing the proof of sufficiency of (38) for (37). *Q.E.D.*

Intuitively, Theorem 2 suggests that even if one does not have information on inputs and outputs, one can still obtain the same information about the scale economies or diseconomies inherited in that technology by using the dual (profit-based) scale elasticity measure defined in (36) or its simplified (and equivalent) version given in (40), provided that the necessary and sufficient condition (38) is satisfied and that the standard regularity conditions of production theory (i)-(v), and differentiability assumptions hold. On the other hand, even if one does not have information on prices or cannot obtain/estimate the profit function (8), but can obtain DDF (6), which only requires input-output data, one can still find the optimal level of scale economies or diseconomies suggested by the profit function of a profit-maximizing agent—by evaluating the scale elasticity measure based on the DDF at the profit-maximizing input-output allocations. Importantly, note that this result does not require assumption that technology set T is convex.

Notably, theorem 1 and 2 together imply that the optimal level of scale economies or diseconomies for a profit-maximizing agent can also be found without knowledge of the directional distance function, just by using scale elasticity measures based on the input oriented or the output oriented Shephard's distance functions, evaluating them at the profit-maximizing input-output allocations.

6. Concluding remarks

In this work we investigated equivalences between various measures of scale elasticity for multi-output-multi-input technologies. We first introduced a scale elasticity measure based on the directional distance function and then derived the necessary and sufficient condition for its equivalence with scale elasticity measures based on the Shephard's distance functions. We also established equivalence relationship between our scale elasticity measure based on the directional distance function and a scale elasticity measure based on the profit function. We proved our results in somewhat classical for economics fashion, using the Lagrangian function and the envelop theorem.

We believe that our result, although theoretical, is valuable for empirical researchers as it provides a testable (necessary and sufficient) conditions that answer when (and only when) the alternative definitions of scale elasticity, primal or dual, yield equivalent conclusions about economies or diseconomies of scale.

A natural extension to this work might involve some Monte Carlo analysis on whether the equivalence we established in this work also holds in the context of various estimators, at least approximately (given some estimation noise) or on average, or at the median, and under what level of noise and what types of conditions on the estimators. Such work is a subject in itself and so is left for future research.

References

- Chambers, R.G., Y. Chung, R., Färe (1996) "Benefit and distance functions," *Journal of Economic Theory* 70:2, pp. 407-419.
- Chambers, R.G., Y. Chung, R., Färe (1998) "Profit, directional distance functions, and Nerlovian efficiency," *Journal of Optimization Theory and Applications* 98, pp. 351-364.
- Chau, N., R. Färe (2011) "Shadow Pricing Market Access: A Trade Benefit Function Approach" *Journal of Economic Theory*, forthcoming.
- Chau, N., R. Färe, and S. Grosskopf (2003) "Trade restrictiveness and efficiency," *International Economic Review* 44(3), pp. 1079-1095.
- Chung, Y.H., Färe, R., Grosskopf, S. (1997) "Productivity and undesirable outputs: a directional distance function approach." *Journal of Environmental Management* 51, pp. 229-240.
- Färe, R., S. Grosskopf, K. Hayes and D. Margaritis (2008) "Estimating Translated Demand Functions," *Journal of Econometrics* 147, pp. 266-274.
- Färe, R., S. Grosskopf and C.A.K. Lovell (1986), "Scale Economies and Duality" *Zeitschrift für Nationalökonomie* 46:2, pp. 175-182.
- Färe, R., S. Grosskopf, D.W. Noh, and W. Weber (2005) "Characterization of a polluting technology: theory and practice," *Journal of Econometrics* 126, pp. 469-492.
- Färe and Karagiannis (2011) "Scale and translation elasticities for directional distance functions," work in progress.
- Färe, R., and D. Primont (1995) *Multi-Output Production and Duality: Theory and Applications*. Kluwer Academic Publishers, Boston.
- Färe, R., Primont, D. (2006) "Directional duality theory," *Economic Theory* 29, pp. 239-247.
- Hanoch, G. (1975) "The Elasticity of Scale and the Shape of Average Costs," *American Economic Review* 65, pp. 492-497.
- Luenberger, D.G., (1992) "Benefit functions and duality," *Journal of Mathematical Economics* 21, pp. 461-481.

- Luenberger, D.G. (1994) “Optimality and the Theory of Value.” *Journal of Economic Theory* 63, pp. 147-69.
- Luenberger, D.G. (1995) *Microeconomic Theory*. McGraw-Hill, Boston.
- Panzar, J.C. and R.D. Willig (1977) “Economies of Scale in Multi-Output Production, *Quarterly Journal of Economics* 91, pp. 481-493.
- Shephard, R.W. (1953) *Cost and production functions*. Princeton, NJ: Princeton University Press.
- Shephard, R.W. (1970) *Theory of cost and production functions*. Princeton, NJ: Princeton University Press.
- Zelenyuk, V. (2011) “A Note on Equivalences in Measuring Returns to Scale in Multi-output-multi-input Technologies” Working Paper Series No. WP05/2011, Centre for Efficiency and Productivity Analysis, School of Economics, University of Queensland, Australia.

NOTES

ⁱ E.g., see Chau et al. (2011), Chung et al. (1997), Chambers et al., (1996, 1998), Färe et al. (2005, 2008), Luenberger (1992, 1994, 1995), to mention just a few.

ⁱⁱ See Chambers et al. (1996, 1998) and Färe and Primont (2006) and related results in Luenberger (1992).

ⁱⁱⁱ It appears that at about the same time, independently, similar definition was used in a paper of Färe and Karagiannis (2011), who sent me their paper in progress immediately after this paper was sent to them.

^{iv} To be precise, note that $D_o(x, \mathbf{0}_M) = 0$ and $D_i(\mathbf{0}_M, x) = +\infty$, $\forall x \geq \mathbf{0}_N$, as well as $D_o(\mathbf{0}_N, y) = +\infty$ and $D_i(y, \mathbf{0}_N) = 0$, $\forall y \geq \mathbf{0}_M$, but these peculiar cases are ruled out from our consideration by the definition of the output and input scale elasticity measures.

^v E.g., see Atkinson and Primont (2002), Atkinson, Cornwell and Honerkamp (2003), O'Donnell and Coelli (2005), Färe, Grosskopf, Hayes and Margaritis (2008) and references cited there in, to mention just a few.

^{vi} An exception is the work in progress of Färe and Karagiannis (2011) who independently and using a different proof, based on duality relationship derived in Chambers et al. (1998) and so, in addition to our assumption, also implicitly assuming convexity of the technology set T , establish the result that $p'y/w'x$ equals $-\nabla'_x D_d(x, y | d_x, d_y)x / \nabla'_y D_d(x, y | d_x, d_y)y$. (Their results got to my attention after this paper was finished.)