Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP07/2012

A Scale Elasticity Measure for Directional Distance Function and its Dual: Theory and DEA Estimation

Valentin Zelenyuk

Date:
November 2012

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: Theory and DEA Estimation[1]

Valentin Zelenyuk*

Abstract

In this paper we focus on scale elasticity measure based on directional distance function for multi-output-multi-input technologies, explore its fundamental properties and show its equivalence 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 scale elasticity measure based on the profit function. Finally, we discuss the estimation issues of the scale elasticity based on the directional distance function via the DEA estimator.

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

JEL Code: D20, D24, L25.

This version: November 7, 2012

Acknowledgement: I would like to thank Robert Chambers, Erwin Diewert, Rolf Färe, Jeff Kline, Bogdan Klishchuk, Antonio Peyrache, Chris O'Donnell, Prasada Rao and Natalya Zelenyuk for fruitful discussions and valuable feedback to this paper. I especially thank anonymous referees of The European Journal of Operational Research for their valuable feedback that helped improving this paper. I remain to be solely responsible for my 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] This is an updated version of the WP06/2011 of CEPA, with the new major additions being section 6 and section 7. Some results from the previous version were refined here and some discovered typos were corrected.

1. Introduction

Since its inception by Chambers, Chung and Färe (1996, 1998), and earlier inspirations 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, especially in analysis of production, productivity, efficiency and economic growth, environmental shadow price estimation, etc.[1] 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), Banker (1984), Banker et al (1984), Färe, Grosskopf and Lovell (1986), Banker and Thrall (1992), Førsund (1996), Golany and Yu (1997), as well as more recent and fundamental works of Fukuyama (2000, 2003), Førsund and Hjalmarsson (2004), Krivonozhko et al. (2004), Hadjicostas and Soteriou (2006, 2010), and Podinovski et al. (2009), to mention just a few. In the present work, we extend some of these works for 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 define 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 arbitrarily, 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 alternative definitions of scale elasticity and, in particular, a 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 related to the profit function. Section 6 discusses the case of technology frontiers with 'kinks' and Section 7 provides details for estimation of the scale elasticities based on the DDF in the DEA context. Finally, Section 8 concludes.

2. Characterizations of Technology

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

$$T\equiv\{(x,y)\in\Re_+^N\times\Re_+^M \quad : \quad y\in\Re_+^M \textit{ is producible from } x\in\Re_+^N\}. \tag{1}$$

By convention in production theory, we admit that technology satisfies 'standard regularity conditions' of production theory such as

(i) "no free lunch" ($(0,y)\notin T$ for any $y\neq 0$),

(ii) "doing nothing is possible" ($(x,0)\in T$ for any $x\in\Re_+^N$),

(iii) the set T is a closed set,

(iv) the output sets of T (defined by $P(x) := \{y : (x, y) \in T\}$, $x \in \Re_+^N$) are bounded for any $x \in \Re_+^N$,

(v) technology set T satisfies 'free disposability' for all inputs and all outputs, i.e., $(x, y) \in T \Rightarrow (x_o, y_o) \in T$, $\forall y_o \leqq y$ and $\forall x \leqq x_o$.

For details of these regularity conditions and resulting properties see Färe and Primont (1995) as well as Chambers et al., (1996, 1998).[2]

In a single output case, a common approach to completely characterize the technology set T is to use the production function $f : \Re_+^N \to \Re_+$ defined as

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

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

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

while the Shephard's (1953) input distance function $D_i : \Re_+^M \times \Re_+^N \to \Re_+ \cup \{+\infty\}$ is defined as

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

Under fairly mild regularity conditions on technology stated above, both functions possess many useful properties, in particular, they both completely characterize technology set T in the sense that

$$D_o(x,y) \le 1 \quad \Leftrightarrow \quad D_i(y,x) \ge 1 \quad \Leftrightarrow \quad (x,y) \in T \tag{5}$$

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. Formally, the directional distance function $D_d : \Re_+^N \times \Re_+^M \to \Re \cup \{-\infty\} \cup \{+\infty\}$ is defined as

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

where $d = (d_x, d_y) \in \Re_+^N \times \Re_+^M \setminus \{0\}$ is a direction vector specified by the researcher. (Note that the notation $d = (d_x, d_y)$ does not imply that d depends on x or y, but only indicates that d_x is a vector in x-space and d_y is a vector in y-space.) Throughout this work, we assume that a suitable direction $d = (d_x, d_y)$ is chosen, in the sense that DDF attains a finite value, implying feasibility of DDF.[3]

Many useful properties of this function were established in the literature (e.g., see Chambers, Chung and Färe (1996, 1998), Briec and Kerstens (2009), etc.). A particularly useful property for us is that, under the standard regularity condition (i)-(v) and a suitable direction, the DDF completely characterizes the technology set *T*, in the sense that

$$D_d(x, y \mid d_x, d_y) \ge 0 \quad \Leftrightarrow \quad (x,y) \in T \,. \tag{7}$$

An important advantage of DDF-based characterization of technology over others is that, under certain conditions, it is dual to the profit function $\pi : \Re_+^M \times \Re_+^N \to \Re \cup \{+\infty\}$,[4] defined as

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

where $w \in \Re_+^N$ is an input price vector and $p \in \Re_+^M$ is an output price vector, corresponding to $x \in \Re_+^N$ and $y \in \Re_+^M$, respectively.

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

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

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

where $f(x)$ is the production function defined in (2) that is assumed to be continuously differentiable and where $\nabla_x^{'} f(x) \equiv (\partial f(x)/\partial x_1, ..., \partial f(x)/\partial x_N)$ 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

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

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

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

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 a 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 of 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 (this approach is also coherent with other works, including Førsund and Hjalmarsson (2004), Podinovski et al. (2009), etc.). 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), the 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.

We 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 (e.g., see Fukuyama, 2003), it seems natural to define it on analogy with the measures defined in (10) and (11) and we will do so here, for a given direction. That is, a measure of scale elasticity based on the DDF, for a given direction (d_x, d_y), is defined as

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

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.[5] An alternative definition may involve also differentiation of the directional vector (e.g., see Fukuyama, 2003), yet it seems natural to keep the direction as a conditioning, outside of differentiation, as we do in what follows.[6]

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 $(0, y)$ or $(x, 0)$) the answer is not straightforward. In the next section we establish one important result about such relationship. As many proofs in economics and in optimization problems, the derivation in the proof of this result is facilitated with the Lagrangian method and the envelope 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

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

If there is an interest in an input-output allocation where $(x, y) \in T \backslash T^{\partial}$, then one must choose a direction along which the inefficiency is to be measured, then find a suitable projection of this input-output allocation onto the frontier for this direction, call it $(\hat{x}, \hat{y}) \in T^{\partial}$, and then measure the scale properties at that projection. Importantly, note that for the technically inefficient points (where $(x, y) \in T \backslash T^{\partial}$), 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 the efficiency measurement, and can even suggest opposite conclusions. Therefore, in general, for input-output allocations where at least one of the distance functions suggests technical inefficiency, one cannot guarantee the equivalence except for some very special cases, and so in the derivation of our results we will focus only on $(x, y) \in T^{\partial}$. Moreover, we will focus on the case when all the distance functions are continuously differentiable at points of interest and will consider a more general case of technology frontiers with 'kinks' later, in sections 6 and 7.

Theorem 1.

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

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

if and only if

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

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

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

This result follows from applying the implicit function theorem to (12), and it requires condition $\nabla'_y D_d(x,y\,|\,d_x,d_y)y \neq 0$ to be valid. Similarly, by using the implicit function theorem, we can rewrite definitions in (10) and (11), respectively, also in a more compact form:

$$e_o(x,y) = -\nabla'_x D_o(x,y)x\,, \tag{16}$$

and

$$e_i(y,x) = -\nabla'_y D_i(y,x)y\,. \tag{17}$$ [7]

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 $e_d(x,y\,|\,d_x,d_y)$ $= e_o(x,y) = 1/e_i(y,x) = 0$, which is not possible since $e_i(y,x) = -\nabla'_y D_i(y,x)y < \infty$. 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 (7), we can rewrite the output distance function as following:

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

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

$$L(\theta, \gamma \mid x, y) = \theta - \gamma(D_d(x, y/\theta \mid d_x, d_y) - 0). \tag{19}$$

Let $\theta^* = \theta(x,y)$ and $\gamma^* = \gamma(x,y)$ be optimal solutions to minimizing (19), then the associated f.o.c. is given by

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

and

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

Now, because θ^* is a solution to minimization of (19), its value must be equal to unity due to the fact that $(x, y) \in T^\partial$, and therefore (20) reduces to

$$\gamma^* \nabla'_y D_d(x, y \mid d_x, d_y) y = -1. \tag{22}$$

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

$$\nabla'_x D_o(x,y) = \nabla'_x L(\theta^*, \gamma^* \mid x, y) = -\gamma^* \nabla'_x D_d(x, y/\theta^* \mid d_x, d_y). \tag{23}$$

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

$$-\nabla'_x D_o(x,y) x = \gamma^* \nabla'_x D_d(x, y \mid d_x, d_y) x. \tag{24}$$

Noting that the l.h.s. of (24) is the output oriented scale elasticity, due to (16), and combining this result with (22), and with our assumption that $\nabla_y^{'} D_d(x, y \mid d_x, d_y) y \neq 0$, we obtain

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

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

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

and so the corresponding Lagrangian function for this optimization problem is

$$L(\theta, \lambda \mid x, y) = \lambda - \delta(D_d(x/\lambda, y \mid d_x, d_y) - 0), \tag{27}$$

Let $\lambda^* = \lambda(x, y)$ and $\delta^* = \delta(x, y)$ be optimal solutions to minimization of (27), then the associated f.o.c. is given by

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

and

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

Similarly as above, because λ^* is a solution to minimizing (27), its value must be equal to unity due to the fact that $(x, y) \in T^\partial$, and therefore (28) reduces to

$$\delta^* \nabla_x^{'} D_d(x, y \mid d_x, d_y) x = -1. \tag{31}$$

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

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

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

$$-\nabla'_y D_i(y,x)x = \delta^* \nabla'_y D_d(x, y \mid d_x, d_y) y. \tag{33}$$

Thus, 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 $\nabla'_x D_d(x, y \mid d_x, d_y)x \neq 0$, we obtain

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

Q.E.D.

This theorem is a generalization of the result on equivalence of input and output oriented scale elasticity measures found in Färe, Grosskopf and Lovell (1986) and Zelenyuk (2011), extending the latter to incorporate the scale elasticity measure based on the directional distance function.[8] Also, note that for strictly positive input output allocations, i.e., $(x, y) \in \Re^{N+M}_{++}$, the condition (14) is immediately simplified into

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

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 gradient of the input distance function and the gradient of the directional distance function, both with respect to output vector, shall *not* be orthogonal to the output vector; while the gradient of the directional distance function with respect to input vector shall not be orthogonal to the input vector. Moreover, as stated in (35), in the special case of measuring at strictly positive input-output combination, the necessary and sufficient condition reduces to the requirement that at least one partial derivative of the input distance function and at least one partial derivative of the directional distance function is non-zero.

On one hand, 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* should induce 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 many other points of interest in practice (e.g., at the average, the 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, our theoretical result attains an empirical importance in modelling multi-output-multi-input technologies, where it became very popular to estimate various distance functions. 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, technical 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 establish a duality relationship between these two alternative measures.

5. Dual Equivalences

Various duality results for DDF have been established in the literature (e.g., see Luenberger (1992), Chambers et al. (1996, 1998), Färe and Primont (2006), Briec and Kerstens (2009)). 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 reciprocal of (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.[9]

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

$$e_{\pi}(p,w) = \left.\frac{\partial \log \theta}{\partial \log \lambda}\right|_{\theta=1,\lambda=1}, \qquad \text{such that} \quad \pi(\theta p, \lambda w) = \pi^{o}, \tag{36}$$

where $\pi^o \in \Re$ which can be set to 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 DMU 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 $e_d(x,y\,|\,d_x,d_y)$. Again, we will focus on the case when all the functions are continuously differentiable at the points of interest and will consider a more general case of technology frontiers with 'kinks' later, in sections 6 and 7.

Theorem 2.

Given definitions (1), (6), (8), (12) and (36), standard regularity conditions of production theory (i)-(v) and assuming that in a neighborhood of a point of interest, $D_d(x,y\,|\,d_x,d_y)$ and $\pi(p,w)$ are continuously differentiable w.r.t. each of their argument, we have:

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

if and only if

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

and where

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

Proof: To prove necessity of (38), assume (37) is true and this would immediately require that $\nabla'_x D_d(x^*, y^* \mid d_x, d_y)x^* \neq 0$. Moreover, using implicit function theorem, we can rewrite (36) as

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

which immediately requires that $p'\nabla_p \pi(p,w) \neq 0$, 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:

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

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

$$L(x,y,\eta \mid p,w,d_x,d_y) = p'y - w'x - \eta(D_d(x,y \mid d_x,d_y) - 0). \tag{42}$$

Let $y^* = y(p,w)$, $x^* = x(p,w)$, and $\eta^* = \eta(p,w)$ be solutions to (41), then the system of equations defined by the associated first order condition is given by

$$\nabla_y L \Big|_{\substack{y=y^* \\ x=x^* \\ \eta=\eta^*}} = p - \eta^* \nabla_y D_d(x^*, y^* \mid d_x, d_y) = 0, \tag{43}$$

and

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

and

$$\nabla_\eta L \Big|_{\substack{y=y^* \\ x=x^* \\ \eta=\eta^*}} = D_d(x^*, y^* \mid d_x, d_y) = 0 \tag{45}$$

Furthermore, rearranging (43) and (44), we get

$$p = \eta^* \nabla_y D_d(x^*, y^* \mid d_x, d_y) \tag{46}$$

and

$$-w = \eta^* \nabla_x D_d(x^*, y^* \mid d_x, d_y)\,, \tag{47}$$

which in turn imply that

$$p' y^* = \eta^* \nabla_y' D_d(x^*, y^* \mid d_x, d_y) y^* \tag{48}$$

and

$$-w' x^* = \eta^* \nabla_x' D_d(x, y \mid d_x, d_y) x^*\,. \tag{49}$$

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

$$\nabla_p \pi(p, w) = \nabla_p L(x^*, y^*, \eta^* \mid p, w, d_x, d_y) = y^*\,. \tag{50}$$

and

$$\nabla_w \pi(p, w) = \nabla_w L(x^*, y^*, \eta^* \mid p, w, d_x, d_y) = -x^*\,. \tag{51}$$

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

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

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

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

Now, combining (48) and (49), and assuming that $\nabla_y' D_d(x, y \mid d_x, d_y) y^* \neq 0$, we get

$$\frac{w' x^*}{p' y^*} = -\frac{\nabla_x' D_d(x, y \mid d_x, d_y) x^*}{\nabla_y' D_d(x^*, y^* \mid d_x, d_y) y^*} = e_d(x^*, y^* \mid d_x, d_y)\,. \tag{54}$$

Finally, combining (53) with (54) we arrive to (37), completing the proof.

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), 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 Theorem 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. Practical matters

The theoretical developments outlined above require standard differentiability of the considered functions. In practice, when dealing with real data, estimations of technology sets of interest is usually done with such methods as the Data Envelopment Analysis (DEA) or the regression-based methods, including Stochastic Frontier Analysis (SFA). The regression-based estimators usually (if not always) presume standard differentiability for the technology-characterizing functions and so an application of results of this paper would be direct there. On the other hand, the DEA-type estimators provide piece-wise-linear approximations of technology sets and this creates some difficulties for the measures based on standard derivatives. The main difficulty comes from the fact that DEA-estimated technology frontiers may (and usually do) have 'kinks' and the standard differentiation methods that we involved in previous sections cannot be applied at these 'kinks'. Note, however, that the results are still applicable everywhere else where the usual derivatives of the involved distance functions would exist. Moreover, the 'kinks' can also be handled via more general differentiation methods, e.g., by using the lower Dini (or left-hand) and the upper Dini (or right hand) derivatives (e.g., see Royden, 1988).

To our knowledge, using Dini-type derivatives has been first suggested and adopted to estimation of scale elasticities in the DEA-type framework in the seminal work of Banker and Thrall (1992). Their paradigm was then further elaborated and refined in Banker et al (1996), Førsund (1996), Golany and Yu (1997), Fukuyama (2000, 2003), Førsund and Hjalmarsson (2004), Krivonozhko et al. (2004), Hadjicostas and Soteriou (2006, 2010), Førsund et al (2007), and Podinovski et al. (2009).

Specifically, Banker and Thrall (1992) introduced the concept of the so-called 'left-hand scale elasticity' and the 'right-hand scale elasticity'. In a nutshell, whether defined for the output or the input orientation, the 'left-hand scale elasticity' and the 'right-hand scale elasticity' measures yield different values at the 'kinks' and whenever they yield the same value (not at the kinks) they coincide with the usual scale elasticity measures of the type we defined in (10) and (11).[2] To briefly outline this approach, let $\nabla_y^- D_i(y,x)$ and $\nabla_y^+ D_i(y,x)$ denote the vectors of the left-hand and the right-hand partial derivatives of $D_i(y,x)$ w.r.t. y, while $\nabla_x^- D_o(x,y)$ and $\nabla_x^+ D_o(x,y)$ denote the vectors of the left-hand and the right-hand partial derivatives of $D_o(x,y)$ w.r.t. x. Then, the 'left-hand scale elasticity' and the 'right-hand scale elasticity' at a point $(x,y) \in T^\partial$, based on the Shephard's input distance function are, respectively, given by

$$e_i^-(x,y) = -\nabla_y^- D_i(y,x) \qquad (55)$$

[2] This was envisioned by Banker and Thrall (1992) and formally proven by Hadjicostas and Soteriou (2006), in particular, see their theorems 3.3 and 3.4 for more details.

and

$$e_i^+(x,y) = -\nabla_y^+ D_i(y,x) \tag{56}$$

while, the 'left-hand scale elasticity' and the 'right-hand scale elasticity' at a point $(x,y) \in T^\partial$, based on the Shephard's input distance function are, respectively, given by

$$e_o^-(x,y) = -\nabla_x^- D_o(x,y) \tag{57}$$

and

$$e_o^+(x,y) = -\nabla_x^+ D_o(x,y). \tag{58}$$

Using the same logic as above, the 'left-hand scale elasticity' and the 'right-hand scale elasticity' at a point $(x,y) \in T^\partial$, based on the directional distance function are, respectively, given by

$$e_d^-(x,y|d_x,d_y) = -\frac{x\prime\nabla_x^- D_d(x,y|d_x,d_y)}{y\prime\nabla_y^- D_d(x,y|d_x,d_y)} \tag{59}$$

$$e_d^+(x,y|d_x,d_y) = -\frac{x\prime\nabla_x^+ D_d(x,y|d_x,d_y)}{y\prime\nabla_y^+ D_d(x,y|d_x,d_y)} \tag{60}$$

where, $\nabla_x^- D_d(x,y|d_x,d_y)$ and $\nabla_x^+ D_d(x,y|d_x,d_y)$ denote the vectors of the left-hand and the right-hand partial derivatives of $D_d(x,y|d_x,d_y)$ w.r.t. x, while $\nabla_y^- D_d(x,y|d_x,d_y)$ and $\nabla_y^+ D_d(x,y|d_x,d_y)$ denote the vectors of the left-hand and the right-hand partial derivatives of $D_d(x,y|d_x,d_y)$ w.r.t. y.

Finally, the 'left-hand scale elasticity' and the 'right-hand scale elasticity' based on the profit function are, respectively, given by

$$e_\pi^-(p,w) = -\frac{w'\nabla_w^- \pi(p,w)}{p'\nabla_p^- \pi(p,w)} \tag{61}$$

and

$$e_\pi^+(p,w) = -\frac{w'\nabla_w^+ \pi(p,w)}{p'\nabla_p^+ \pi(p,w)} \tag{62}$$

where $\nabla_p^- \pi(p,w)$ and $\nabla_p^+ \pi(p,w)$ denote the vectors of the left-hand and the right-hand partial derivatives of $\pi(p,w)$ w.r.t. p, while $\nabla_w^- \pi(p,w)$ and $\nabla_w^+ \pi(p,w)$ denote the vectors of the left-hand and the right-hand partial derivatives of $\pi(p,w)$ w.r.t. w.

The theoretical results and their proofs that we derived in previous sections for the case of standard derivatives can now be re-produced here with replacement of the standard derivatives with the left-hand and the right-hand derivatives.

Specifically, under the standard regularity conditions of production theory (i)-(v) and assuming that in a neighborhood of a point of interest $(x,y) \in T^\partial$ and where the functions $\nabla_x^- D_d(x,y|d_x,d_y)$, $\nabla_y^- D_d(x,y|d_x,d_y)$, $\nabla_x^+ D_d(x,y|d_x,d_y)$, $\nabla_y^+ D_d(x,y|d_x,d_y)$, $\nabla_y^- D_i(y,x)$, $\nabla_y^+ D_i(y,x)$, $\nabla_x^- D_o(x,y)$ and $\nabla_x^+ D_o(x,y)$ exist, we have[10]

$$e_d^-(x,y|d_x,d_y) = e_o^-(x,y) = 1/e_i^-(x,y) \tag{63}$$

if and only if

$$y\nabla_y^- D_i(y,x) \neq 0 \text{ and } y\nabla_y^- D_d(x,y|d_x,d_y) \neq 0 \text{ and } x\nabla_x^- D_d(x,y|d_x,d_y) \neq 0, \tag{64}$$

while

$$e_d^+(x,y|d_x,d_y) = e_o^+(x,y) = 1/e_i^+(x,y) \tag{65}$$

if and only if

$$y\nabla_y^+ D_i(y,x) \neq 0 \text{ and } y\nabla_y^+ D_d(x,y|d_x,d_y) \neq 0 \text{ and } x\nabla_x^+ D_d(x,y|d_x,d_y) \neq 0. \tag{66}$$

Moreover, given the standard regularity conditions of production theory (i)-(v) and assuming that $\nabla_x^- D_d(x,y|d_x,d_y)$, $\nabla_y^- D_d(x,y|d_x,d_y)$, $\nabla_x^+ D_d(x,y|d_x,d_y)$, $\nabla_y^+ D_d(x,y|d_x,d_y)$ and $\nabla_p^- \pi(p,w)$, $\nabla_w^- \pi(p,w)$, $\nabla_p^+ \pi(p,w)$ and $\nabla_w^+ \pi(p,w)$ exist, we have

$$e_\pi^+(p,w) = e_d^+(x^*,y^*|d_x,d_y) \tag{67}$$

if and only if

$$y^{*\prime}\nabla_y^{+} D_d\left(x^*, y^* \middle| d_x, d_y\right) \neq 0 \quad \text{and} \quad p'\nabla_p^{+}\pi(p, w) \neq 0 \tag{68}$$

and

$$e_\pi^{-}(p, w) = e_d^{-}\left(x^*, y^* \middle| d_x, d_y\right) \tag{69}$$

if and only if

$$y^{*\prime}\nabla_y^{-} D_d\left(x^*, y^* \middle| d_x, d_y\right) \neq 0 \quad \text{and} \quad p'\nabla_p^{-}\pi(p, w) \neq 0 \tag{70}$$

and where

$$(x^*, y^*) = \arg\max_{x,y} \left\{p'y - w'x : (x, y) \in \hat{T}_{VRS}\right\}. \tag{71}$$

In the next section we give more details on the DEA formulations to obtain the estimates of scale elasticities based on the directional distance function.

7. The DEA Formulations[11]

To facilitate our further discussion, recall that given a sample of input-output data, $\{(x_j, y_j) : j = 1, \ldots n\}$, on some n DMUs, the DEA estimator of the true but unobserved technology set, allowing for variable returns to scale (VRS) technology, is given by

$$\hat{T}_{VRS} = \left\{(x, y) : \ x \geq \textstyle\sum_{j=1}^{n} z_j x_j \, , y \leq \textstyle\sum_{j=1}^{n} z_j y_j \, , \ \textstyle\sum_{j=1}^{n} z_j = 1, \ z_j \geq 0, \ j = 1, \ldots, n\right\}. \tag{72}$$

Intuitively, (72) gives the smallest convex, free disposal hull that fits the data $\{(x_j, y_j) : j = 1, \ldots n\}$. Now, to obtain an estimate of the scale elasticity based on the directional distance function via the DEA estimator (72), one can adopt the following procedure:

Step 1. Check if condition $(x_k, y_k) \in T^{\partial}$ is satisfied for the points of interest w.r.t. the DEA-estimated technology. To do this, obtain estimates of the involved distance functions at the point of interest $(x_k, y_k) \in T$ and check if these estimates are all indicating full technical efficiency. Specifically, the DEA estimate of the reciprocal of the input oriented Shephard's distance function, for a particular DMU k, can be obtained by solving the following linear program (LP), hereafter LP1:

$$\hat{\lambda}_k := \min \lambda_k$$

$$s.t. \tag{73}$$

$$\lambda_k x_k \geq \sum_{j=1}^{n} z_j x_j,$$

$$y_k \leq \sum_{j=1}^{n} z_j y_j,$$

$$\lambda_k \geq 0, \quad \sum_{j=1}^{n} z_j = 1, \quad z_j \geq 0, \quad j = 1, \dots, n$$

The DEA estimate of the reciprocal of the output oriented Shephard's distance function, for a particular DMU k, can be obtained by solving the following LP (hereafter LP2):

$$\hat{\theta}_k := \max \theta_k$$

$$s.t. \tag{74}$$

$$x_k \geq \sum_{j=1}^{n} z_j x_j,$$

$$\theta_k y_k \leq \sum_{j=1}^{n} z_j y_j,$$

$$\theta_k \geq 0, \quad \sum_{j=1}^{n} z_j = 1, \quad z_j \geq 0, \quad j = 1, \dots, n$$

Similarly, the DEA estimate of the directional distance function, for a particular DMU k, can be obtained by solving the following LP (hereafter LP3):

$$\hat{\beta}_k := \max \beta_k$$

$$s.t. \tag{75}$$

$$x_k - \beta_k d_x \geq \sum_{j=1}^{n} z_j x_j,$$

$$y_k + \beta_k d_y \leq \sum_{j=1}^{n} z_j y_j,$$

$$\beta_k \geq 0, \quad \sum_{j=1}^{n} z_j = 1, \quad z_j \geq 0, \quad j = 1, \dots, n$$

If (73)-(75) yield $\hat{\lambda}_k = \hat{\theta}_k = 1$ and $\hat{\beta}_k = 0$, (i.e., indicate full efficiency of (x_k, y_k)) then step 2 is taken. If at least one of them is violated, then the three elasticity measures are not guaranteed to give the same results. In such a situation, it might be of interest to obtain scale elasticity estimates at a projection of the observation of interest (x_k, y_k) onto some point $(\hat{x}_k, \hat{y}_k)$ on the DEA frontier where $\hat{\lambda}_k = \hat{\theta}_k = 1$ and $\hat{\beta}_k = 0$ hold. For example, if $\hat{\lambda}_k < 1$, while $\hat{\theta}_k = 1$ and $\hat{\beta}_k = 0$, then a suitable projection is given by $(\hat{x}_k, \hat{y}_k) \coloneqq (x_k \hat{\lambda}_k, y_k)$, or if $\hat{\theta}_k > 1$, while $\hat{\lambda}_k = 1$ and $\hat{\beta}_k = 0$, then a suitable projection is $(\hat{x}_k, \hat{y}_k) \coloneqq (x_k, \hat{\theta}_k y_k)$, or if $\hat{\lambda}_k = \hat{\theta}_k = 1$ while $\hat{\beta}_k > 0$ then a suitable projection is $(\hat{x}_k, \hat{y}_k) \coloneqq (x_k - \hat{\beta}_k d_x, y_k + \hat{\beta}_k d_y)$.

It is also possible (and often happens in practice) that $\hat{\lambda}_k < 1$, $\hat{\theta}_k > 1$ and $\hat{\beta}_k > 0$, i.e., the DMU k is in the interior of $\hat{T}_{VRS}$ and so is not technically efficient w.r.t. any of the three distance functions. In such cases, there is ambiguity about the choice of direction for a suitable projection, in the sense that conclusions about the returns to scale may heavily depend on the choice of the direction. Moreover, note that projection based on only one direction (input or output or along the direction $(-d_x, d_y)$) may not be enough to ensure that $\hat{\lambda}_k = \hat{\theta}_k = 1$ and $\hat{\beta}_k = 0$ at the projected point. Indeed, after projecting, say, along $(-d_x, d_y)$ one might also need to project radially in the input (or the output) space or vice versa. Furthermore, note that different order of projections may yield quite different conclusions about the returns to scale—not only quantitatively but also qualitatively different. So, evaluation of scale elasticity for DMUs that are inefficient according to all of the distance functions involved is a moot point and, perhaps, shall not be pursued.

Step 2. Once it is ensured that for a DMU k, we have $\hat{\lambda}_k = \hat{\theta}_k = 1$ and $\hat{\beta}_k = 0$ at a point of interest $(\hat{x}_k, \hat{y}_k)$ (which could be an original observation or its projection onto the estimated frontier), the Banker-Thrall-type DEA-estimates can be obtained. Due to the equivalence of the scale elasticity measures based on the DDF and those based on the Shephard's distance functions for a general case established above, one can simply focus on Shephard's distance functions— the left-hand and the right-hand scale elasticities for input-oriented case, which we denote as $\hat{e}_i^-(\hat{x}_k, \hat{y}_k)$ and $\hat{e}_i^+(\hat{x}_k, \hat{y}_k)$ and those for the output oriented case, which we denote as $\hat{e}_o^-(\hat{x}_k, \hat{y}_k)$ and $\hat{e}_o^+(\hat{x}_k, \hat{y}_k)$. For these measures to be well-defined at a point of interest $(\hat{x}_k, \hat{y}_k) \in \hat{T}_{VRS}$, additional regularity assumptions on the DEA frontier at these points are required. Specifically, we need:

A1. $(\hat{x}_k, \hat{y}_k) \in \hat{T}_{VRS} \Rightarrow (\delta_1 \hat{x}_k, \delta_2 \hat{y}_k) \in \hat{T}_{VRS}$, for some $\delta_1, \delta_2 \in (1, \infty)$.

A2. $(\hat{x}_k, \hat{y}_k) \in \hat{T}_{VRS} \Rightarrow (\delta_3 \hat{x}_k, \delta_4 \hat{y}_k) \in \hat{T}_{VRS}$, for some $\delta_3, \delta_4 \in (0,1)$.

In words, for a point of interest $(\hat{x}_k, \hat{y}_k)$, adding A1 to condition $\hat{\lambda}_k = \hat{\theta}_k = 1$ and $\hat{\beta}_k = 0$ ensures existence of the right-hand input oriented elasticity, while adding A2 to condition $\hat{\lambda}_k = \hat{\theta}_k = 1$ and $\hat{\beta}_k = 0$ ensures existence of the left-hand output oriented elasticity. Note that these are local assumptions, in the sense that they need to be ensured only at a point of estimation of elasticity (also, see Hadjicostas and Soteriou (2006) for similar conditions and discussions).

If A1 holds, then the Banker-Thrall-type DEA-estimates of the left-hand and right-hand input oriented scale elasticity can be found, respectively, as

$$\hat{e}_i^-(\hat{x}_k, \hat{y}_k) = 1 - u_k^- \tag{76}$$

and

$$\hat{e}_i^+(\hat{x}_k, \hat{y}_k) = 1 - u_k^+ \tag{77}$$

where u_0^- is obtained from the following LP (hereafter LP4):

$$\begin{aligned} & u_k^- := \max \ u_k \\ s.t. \quad & \\ & \sum_{m=1}^{M} \hat{y}_{0m} \mu_m + u_k = 1 \end{aligned} \tag{78}$$

$$\sum_{m=1}^{M} y_{jm}\mu_m - \sum_{i=1}^{N} x_{ji}\xi_i + u_k \leq 0, \;\; j = 1, \dots, n$$

$$\sum_{i=1}^{N} \hat{x}_{ki}\xi_i = 1,$$

$$\xi_i \geq 0, \qquad i = 1, \dots, N, \qquad m = 1, \dots, M, \qquad u_k \text{ free}$$

while u_0^+ is obtained from the following LP (hereafter LP5):

$$u_k^+ \coloneqq \min \;\; u_k \tag{79}$$

subject to the same constraints as in LP4.

On the other hand, if A2 holds, then the Banker-Thrall-type DEA-estimates of the left-hand and right-hand output oriented scale elasticity can be found, respectively, as

$$\hat{e}_o^-(\hat{x}_k, \hat{y}_k) = 1 - v_k^- \tag{80}$$

and

$$\hat{e}_o^+(\hat{x}_k, \hat{y}_k) = 1 - v_k^+ \tag{81}$$

where v_k^- is obtained from the following LP (hereafter LP6):

$$v_k^- \coloneqq \min \;\; v_k$$

$$s.t. \tag{82}$$

$$\sum_{i=1}^{N} \hat{x}_{ki}\xi_i + v_k = 1$$

$$\sum_{i=1}^{N} x_{ji}\xi_i - \sum_{m=1}^{M} y_{jm}\mu_m + v_k \geq 0, \qquad j = 1, \dots, n$$

$$\sum_{m=1}^{M} \hat{y}_{km}\mu_m = 1,$$

$$\xi_i \geq 0, \qquad i = 1, \dots, N, \qquad m = 1, \dots, M, \qquad v_k \text{ free}$$

while v_k^+ is obtained from the following LP (hereafter LP7):

$$v_k^+ \coloneqq \max \;\; v_k \tag{83}$$

subject to the same constraints as LP6.

It might be worth pausing and noting here that if A1 does not hold, then $u_k^+ = -\infty$ and so $\hat{e}_i^+(\hat{x}_k, \hat{y}_k)$ would be ∞ or undefined. Similarly, if A2 does not hold, then $v_k^- = -\infty$ and so $\hat{e}_o^+(\hat{x}_k, \hat{y}_k)$ would be ∞ or undefined.

Note that, because of (63) and (65), as long as (64) and (66) hold, one would only need to compute two of those measures, e.g., the input oriented 'right-hand scale elasticity' $\hat{e}_i^+(\hat{x}_k, \hat{y}_k)$ and the output oriented 'left-hand scale elasticity' $\hat{e}_o^-(\hat{x}_k, \hat{y}_k)$ or, alternatively, $\hat{e}_i^-(\hat{x}_k, \hat{y}_k)$ and $\hat{e}_o^+(\hat{x}_k, \hat{y}_k)$. For the points when these measures are not equivalent but are finite numbers, it is useful to present both estimates, for the reader to see the ambiguity and, perhaps, qualitatively different conclusions, which can be hidden if an average is presented. Finally, for the points when one is finite and the other is undefined (or ∞), it would also be useful to present both estimates, for the reader's convenience.

8. Concluding remarks

In this work we investigated equivalences between various measures of scale elasticity for multi-output-multi-input technologies. We focused on the scale elasticity measure based on the directional distance function and derived the necessary and sufficient condition for its equivalence with scale elasticity measures based on the Shephard's distance functions. We also established a dual equivalence: a relationship between the scale elasticity measure based on the directional distance function and the scale elasticity measure based on the profit function. We proved our results using the Lagrangian function and the envelope theorem. We also discussed the case of measuring scale elasticities in the case of technology frontiers with 'kinks' (e.g., piece-wise linear technologies) and provided details for DEA estimation of the scale elasticities based on the directional distance function based on the Banker-Thrall paradigm.

Our result, although theoretical, is valuable for empirical researchers as it provides 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 the present work would be a development of the statistical procedures for obtaining estimates of standard errors and confidence intervals for the scale elasticity measures, as well as for various statistical tests for them. Such work is a subject in itself and so is left for future research.

References

Banker, R.D. (1984) "Estimating most productive scale size using data envelopment analysis," *European Journal of Operational Research* 17, 35-44.

Banker, R., A. Charnes, W.W. Cooper (1984), "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," Management Science 30:9, 1078-92.

Banker, R.D., and Thrall, R.M. (1992) "Estimation of returns to scale using Data Envelopment Analysis," *European Journal of Operational Research* 62, 74-84.

Banker, R.D., Bardhan, I., Cooper, W.W. (1996) "A note on returns to scale in DEA" *European Journal of Operational Research* 88, 583-585.

Briec. W and K. Kerstens (2009) "Infeasibility and Directional Distance Functions with Application to the Determinateness of the Luenberger Productivity Indicator," *Journal of Optimization Theory and Applications* 141, 55-73.

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.

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.

Førsund, F.R. (1996) "On the calculation of the scale elasticity in DEA models," *Journal of Productivity Analysis* 7, 283-302.

Førsund, F.R. and L. Hjalmarsson (2004) "Calculating scale elasticity in DEA models," *Journal of the Operational Research Society* 55, 1023–1038.

Førsund, F. R., and L. Hjalmarsson (1979) "Generalised Farrell measures of efficiency: An application to milk processing in Swedish dairy plants," *The Economic Journal* 89, 294–315.

Fukuyama, H. (2000) "Returns to scale and scale elasticity in data envelopment analysis." *European Journal of Operational Research* 125, 93–112.

Fukuyama, H. (2003) "Scale Characterizations in a DEA Directional Technology Distance Function Framework," *European Journal of Operational Research* 144, 108-127.

Golany, B. and Yu, G. (1997) "Estimating returns to scale in DEA," *European Journal of Operational Research* 103, 28-37.

Hadjicostas, P., and Soteriou, A.C. (2006) "One-sided elasticities and technical efficiency in multi-output production: a theoretical framework." *European Journal of Operational Research* 168, 425-449.

Hadjicostas, P., and Soteriou, A.C. (2010) "Different orders of one-sided elasticities in multi-ouput production," *Journal of Productivity Analysis* 33, 147-167. Erratum (2010), 34, 181.

Hanoch, G. (1975) "The Elasticity of Scale and the Shape of Average Costs," *American Economic Review* 65, pp. 492-497.

Krivonozhko, V.E., Utkin, O.B., Volodin, A.V., Sablin, I.A., and Patrin, M. (2004), "Constructions of economic functions and calculations of marginal rates in DEA using parametric optimization methods," *Journal of the Operational Research Society* 55, 1049-1058.

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, 481-493.

Podinovski, V.V., Førsund, F.R., and Krivonozhko, V.E. (2009) "A simple derivation of scale elasticities in data envelopment analysis," *European Journal of Operational Research* 197, 149-153.

Royden H.L. (1988) Real Analysis, 3rd Ed., Macmillan, New York.

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

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

[2] It might be worth noting that not all of these conditions are required to be satisfied to establish the main results of the paper and we admit these conditions here to be coherent with axiomatic framework that is frequently used in economic theory of production. Specifically, these regularity conditions ensure desirable (from economic theory perspectives) properties of the distance functions employed in this paper, which were derived in various works we cited and relied upon here. Further generalizations or extensions are possible.

[3] See Briec and Kerstens (2009) for thorough theoretical discussions on this issue.

[4] See Chambers et al. (1996, 1998) and Färe and Primont (2006) and related results in Luenberger (1992, 1994, 1995).

[5] Similar definition appeared in the work in progress by Färe and Karagiannis (2011).

[6] It is worthwhile to note here that the definitions of scale elasticity of the type given in (10), (11) and (12) (as well as the one defined later, in (36)) relate the vector of inputs to the vector of outputs implicitly, and so rely on the implicit function theorem and assumptions required by it.

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

[8] For related results, also see Hanoch (1975), Panzar and Willig (1977), Färe, Grosskopf and Lovell (1986), Banker and Thrall (1992), Golany and Yu (1997), as well as more recent works of

Krivonozhko et al. (2004), Podinovski et al. (2009) and most extensively in Hadjicostas and Soteriou (2006, 2010).

[9] An exception is the working paper of Färe and Karagiannis (2011) who established some similar results using alternative strategy of proof. They used duality relationship derived in Chambers et al. (1998) and required stricter assumptions (convexity of T) than we do in our work. (Their work got to my attention after similar and more general results of this paper were established and shared with them.)

[10] A related result (for equivalence of input and output oriented elasticity measures) was formally established by Hadjicostas and Soteriou (2006) for general convex technologies and their proof can also be adopted to provide a proof for the elasticity measures based on the directional distance function. The equivalence of the input and output oriented scale elasticity measures also is given (without a proof) in Podinovski et al (2009).

[11] We thank the editor and anonymous referees for inspiring this section.