A New Metric for Scale Elasticity in Data Envelopment Analysis

M. Hasannasab, I. Roshdi, D. Margaritis, P. Rouse[1]

Abstract

Robust measurement of scale elasticity (SE) in data envelopment analysis (DEA) models remains elusive, primarily reflecting the computational challenges brought about by the piecewise linear nature of the DEA technology. SE is meaningfully defined only at frontier points or at the projection of interior points to the frontier but not for the inefficient unit itself. A long held issue of concern is that returns to scale (RTS) are not uniquely determined for efficient units since they may be located on vertices or on ridges of the efficient fronier. Thus the multiplier of the convexity constraint defining scale elasticity can take on multiple values with existing methods providing different ways to estimate intervals for SE values in which RTS determination may be ambiguous. In this paper, we propose a linear programming (LP) model providing a unique measure of SE based on a simple proposition of closeness to most productive scale size (MPSS). The model is non-oriented, and the result not only provides the SE measure but also allows RTS to be classified. Furthermore, the model yields feasible consistent results using different solvers, while the same can not be said for the results of other methods. The model was tested using a data set of 95 banks across 11 years and compared with the interval approaches commonly used in the literature.

Keywords: Production economics; Scale elasticity; Returns to scale; Most productive scale size; Data envelopment analysis.

1 Introduction

During the past thirty years, researchers in production economics and operations research have made numerous attempts to characterise the scale properties of the production technology based on qualitative and quantitative measures of returns to scale. Panzar and Willig (1977), Starrett (1977), Färe, Grosskopf and Lovell (1985), Banker and Thrall (1992), Førsund (1996), Sueyoshi (1999), Fukuyama (2000), Podinovski, Førsund and Krivonozhko (2009), Balk, Färe and Karagiannis (2015), to name but a few, have made important contributions to the theory and measurement of RTS and its associated scale elasticity metric. An important feature of these approaches is that they are directly applicable to –multiple-output technologies. Scale elasticity is thus defined in terms of distance functions and is computed from Farrell efficiency measures and the multiplier (shadow value) of the convexity constraint.

However existing methods of calculating scale elasticity for frontier points based on non-parametric piecewise linear technologies suffer from inherent ambiguities in choosing a single

[1] **Correspondence:**MHasannasab, The University of Auckland Business School, 12 Grafton Rd, Auckland, New Zealand. Phone: +64 221511623 Email: Maryam.hasannasab@auckland.ac.nz

candidate as the final SE value, e.g., SE may not be uniquely determined between an upper or lower bound or a particular value from the SE interval unambiguously chosen since the efficient point may lie on several efficient facets. Further complications arise as the chosen value may be unrealistic, very sensitive to even small changes in input or output values as well as sensitive to the orientation of the DEA model .

In this paper, we develop an approach that leads to a new metric for SE calculated as the distance from the most productive scale size frontier. Calculating SE against MPSS gives a better economic meaning to the concept of SE since MPSS corresponds to the point on the efficient frontier with the maximum average productivity for the given input-output vectors (see Banker, 1984). Our approach also overcomes the limitations of existing methods in the sense that it offers a single LP model specifying a single meaningful numerical value for SE which provides the RTS characterisation directly, it is not sensitive to changes in inputs and outputs, and is consistent across the solvers we use to solve the proposed model.

The classical method of calculating SE in production economics is related to the direct derivatives of the production transformation function and hence cannot be directly applied in DEA since the efficient frontier of DEA technology is piecewise linear and not everywhere differentiable. Using the dual formulation of the LP program, Banker, Charnes and Cooper(1984) and Banker and Thrall (1992) developed a method based on the observation that the partial derivatives of the transformation function are characterised by the normal vector of the supporting hyperplane of the technology set where the efficient frontier is smooth, and, this can be obtained by the optimal solution of an LP model. Where the technology function is not smooth, more likely for real world data, the investigation of alternative optimal solutions is required. Thus additional LP models are solved to determine left-hand and right-hand SE (e.g., Golany and Yu, 1997) or the lower and upper bound and consequently an interval containing all possible values for SE (e.g., Banker, 1984).

The formulation for calculating intervals for SE in DEA literature also differs depending on the type of the DEA model. These include the input oriented and output oriented SE intervals developed by Färe, Grosskopf and Lovell (1988), and the directional technology scale elasticity measures developed by Fukuyama (2003) and Balk et al. (2015). See (Zelenyuk, 2013) for more details on the relationship between these methods.

Other approaches look at how to measure SE for production points inside the frontier making use of directional projection points (Førsund and Hjalmarsson, 2004), using reference points as projections (Banker, Cooper, Seiford, Thrall and Zhu, 2004) or adjusting SE (Podinovski et al., 2009).[2] These methods relate to our approach since their bottom line is the efficient frontier albeit they encounter some of the limitations of the frontier approaches mentioned above. Similar to previous studies our focus is on variable returns-to-scale (VRS) technologies,

[2] Note that SE measurement is meaningful only for units on the frontier either efficient points or projections of inefficient points thereof recognising however that points inside the frontier do play a role in conveying useful information about RTS.

specifically on the frontier of the production possibilities set ($\mathcal{T}_{\text{VRS}}$) that can be characterised by its supporting hyperplanes. Thus, our proposed model has no dependency on the choice of DEA model orientation or directional distance function model and can be used without prior knowledge of the efficient, and the strong frontier or interior points of $\mathcal{T}_{\text{VRS}}$.

The paper is organized as follows. In Sections 2, for the convenience of the reader, we introduce the relevant technical material on scale elasticity, a summary of existing methods and their drawbacks. Section 3 describes our approach and produces the LP formulation for calculating SE. Section 4 presents an example comparing the proposed model solution with existing methods based on two different groups of solvers.

2. Scale Elasticity and DEA

2.1 Technology characterization and neoclassical scale elasticity

Assume there is a production process transforming a vector of *m* inputs, $\mathbf{x} = (x_1, \dots, x_m)^T \in \mathbb{R}^m_{\geq 0}$, into a vector of *s* outputs, $\mathbf{y} = (y_1, \dots, y_s)^T \in \mathbb{R}^s_{\geq 0}$. [3] A convenient way to describe the production technology of this multi-input, multi-output process is to use the *technology set* that comprises all possible input-output vectors as

$$\mathcal{T} = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{m+s}_{\geq 0} | \quad \mathbf{y} \;\; can\ be\ produced\ \ by \quad \mathbf{x}\}. \tag{1}$$

By imposing some regularity conditions on $\mathcal{T}$ (see Panzar and Willig, 1977), there exists a differentiable implicit function $F(\mathbf{x}, \mathbf{y})$, called a *production transformation function* or *technical transformation function*, that alternatively characterizes the technology, i.e.[4]

$$F(\mathbf{x}, \mathbf{y}) \geq 0 \Leftrightarrow (\mathbf{x}, \mathbf{y}) \in \mathcal{T}. \tag{2}$$

For a single-output production process, this function reduces to the familiar notion of the *production function* which explicitly can be written as $y = f(\mathbf{x})$ where y is the maximal producible value for every $\mathbf{x}$.

The standard measure of SE at every frontier point $(\mathbf{x}, \mathbf{y})$ of $\mathcal{T}$ (i.e. $F(\mathbf{x}, \mathbf{y}) = 0$) is given by Panzar and Willig (1977) as the maximum equiproportionate change in all outputs subject to an equiproportionate increase in all inputs. They further demonstrate that SE can be directly calculated from the derivatives of the production transformation function as[5]

---

[3] The superscript T stands for a vector transpose and bold letter represents a vector.

[4] The production technology can be equivalently represented in terms of input and output sets. (For more details see e.g., Shepherd, 1970; Färe et al., 1985)

[5] Exploiting the interrelationships between the transformation function and Farrell's input and output distance functions, Färe et al. (1985) derived analogous formula for measuring SE in terms of input and output distance functions.

$$\mathcal{E}(\mathbf{x},\mathbf{y}) = -\frac{\sum_{i=1}^{m} \frac{\partial F(\mathbf{x},\mathbf{y})}{\partial x_i} x_i}{\sum_{r=1}^{s} \frac{\partial F(\mathbf{x},\mathbf{y})}{\partial y_r} y_r}. \tag{3}$$

It can be shown that[6] $\mathcal{E}(\mathbf{x},\mathbf{y})$ is the marginal change in outputs scaling factor allowed by a marginal change in the inputs scaling factor over the average ratio of these factors, which reduces to the ratio of marginal productivity to average productivity for the simple single-input, single-output case. For a geometrical interpretation of SE see Cooper, Seifordand Tone(2007).

2.2 SE Measurement in DEA

2.2.1 DEA Piecewise Linear Technology and SE Characterisation

Every DEA model is based on an underlying empirical technology set, constructed from a set of actual observations along with some standard assumptions about the production technology. Following Banker et al. (1984) the production possibilities set under the variable returns to scale DEA technology is formulated by imposing convexity, monotonicity, and minimum extrapolation postulates as

$$\mathcal{T}_{\text{VRS}} = \{(\mathbf{x},\mathbf{y}) \in \mathbb{R}_{\geq 0}^{m} \times \mathbb{R}_{\geq 0}^{s} \colon \mathbf{x} \geq \sum_{j=1}^{n} \lambda^j \mathbf{x}^j \ , \ \mathbf{y} \leq \sum_{j=1}^{n} \lambda^j \mathbf{y}^j , \ \sum_{j=1}^{n} \lambda_j = 1\}, \tag{4}$$

in which $\left(\mathbf{x}^j,\mathbf{y}^j\right)^T$, $j = 1,\dots,n$, denote the input-output vector of the jth observation, generically referred as decision-making units (DMUs) in the DEA terminology.

It is well known that $\mathcal{T}_{\text{VRS}}$ is a polyhedral set and (4) exhibits its *primal* representation. The *dual* representation of this polyhedron is also possible in which $\mathcal{T}_{\text{VRS}}$ is described in terms of its supporting hyperplanes (See, e.g., Briec and Leleu, 2003). Every supporting hyperplane of this special polyhedron, $\mathcal{T}_{\text{VRS}}$, is associated with an equation of the form $\bar{\mathbf{u}}^T y - \bar{\mathbf{v}}^T x + \bar{u}_o = 0$ where $(\bar{\mathbf{u}},\bar{\mathbf{v}}) \geq 0$. In this regard, we can define three different concepts, efficient, strong *efficient* and *inefficient (interior)* point, for each production point, $(\bar{\mathbf{x}},\bar{\mathbf{y}})$, based on its location in $\mathcal{T}_{\text{VRS}}$.

Definition 1 *A production point* $(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}}) \in \mathcal{T}_{VRS}$[7] *is called an efficient point, if it lies on a supporting hyperplane of* $\mathcal{T}_{VRS}$, *i.e., if there exists a vector* $(\bar{\boldsymbol{u}},\bar{\boldsymbol{v}}) \geq 0$ *along with a scalar* $\bar{u}_o$ *such that*

$$\bar{\mathbf{u}}^T \mathbf{y}^j - \bar{\mathbf{v}}^T \mathbf{x}^j + \bar{u}_o \leq 0 \ \text{ for all } \ j = 1,\dots,n \quad \text{and} \quad \bar{\mathbf{u}}^T \bar{\mathbf{y}} - \bar{\mathbf{v}}^T \bar{\mathbf{x}} + \bar{u}_o = 0. \tag{5}$$

---

[6] See, e.g., Førsund and Hjalmarsson (2004) for more details.

[7] In this definition, $(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}}) \in \mathcal{T}_{VRS}$, can be an observed DMU or a projection of an observed DMU.

The efficient point $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ *is called a strong efficient point if there is a vector satisfying (5) in* which all the input-output multipliers are strictly positive, i.e., $(\bar{\mathbf{u}}, \bar{\mathbf{v}}) > 0$. Finally, $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is called an *inefficient* point of $\mathcal{T}_{\mathrm{VRS}}$, if there is no vector $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o) \neq$ satisfying (5).[8]

There is also a relationship between the normal vectors of *supporting* hyperplanes of $\mathcal{T}_{\mathrm{VRS}}$ at an efficient point and the optimal solutions of the input and output-oriented multiplier form of VRS models, presented in (6) below. These models were introduced by Banker (1980) to evaluate the relative efficiency of a given production point $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ operating under VRS:

$$
\begin{array}{lll}
\max\ \mathbf{u}^T\bar{\mathbf{y}} + u_o & & (IM) \\
\quad s.t.\, \mathbf{u}^T\mathbf{y}^j - \mathbf{v}^T\mathbf{x}^j + u_o & \leq 0, & \\
\quad \mathbf{v}^T\bar{\mathbf{x}} & = 1, & \\
\quad \mathbf{u} \geq 0, \mathbf{v} & \geq 0, & \\
\quad u_o\ unrestricted, & &
\end{array}
\qquad
\begin{array}{lll}
\min\ \mathbf{v}^T\bar{\mathbf{x}} - u_o & & (OM) \\
\quad s.t.\, \mathbf{u}^T\mathbf{y}^j - \mathbf{v}^T\mathbf{x}^j + u_o & \leq 0, & \\
\quad \mathbf{u}^T\bar{\mathbf{y}} & = 1, & \\
\quad \mathbf{u} \geq 0, \mathbf{v} & \geq 0, & \\
\quad u_o\ unrestricted & &
\end{array}
$$

In the following theorem, we show that the normal vector of every supporting hyperplane of $\mathcal{T}_{\mathrm{VRS}}$ at a given efficient point can be derived from the optimal solutions of these models.

Theorem 1 *The following statements are true for a given* efficient *point,* $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$*:*

• The normal vector of every supporting hyperplane of $\mathcal{T}_{\mathrm{VRS}}$ at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ gives an optimal solution of either Model (IM), or (OM), or both.

• Every optimal solution of Model (IM) or (OM) corresponds to a supporting hyperplane of $\mathcal{T}_{\mathrm{VRS}}$ at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$.

Proof. The proof is straightforward and follows from the proof of Theorem 5.2 in Cooper et al. (2007) and Remark 1 below.

Remark 1 Theorem 1 *is a slightly generalized version of Theorem 5.2 stated in* Cooper et al. (2007). *They indicate that the normalized coefficient of any supporting hyperplane of* $\mathcal{T}_{VRS}$ *at* $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ *gives an optimal solution of (IM). This is not generally true and there are some special cases where the proposition does not hold. Let* $(\bar{\boldsymbol{u}}, -\bar{\boldsymbol{v}})$ *be the normal vector of a supporting hyperplane of* $\mathcal{T}_{VRS}$ *at* $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$*. It can be seen that if* $\bar{\boldsymbol{v}}^T\bar{\boldsymbol{x}} = 0$*, then* $(\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$ *does not correspond to any optimal solution of Model (IM) and, analogously, if* $\bar{\boldsymbol{u}}^T\bar{\boldsymbol{y}} = 0$*, then* $(\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}})$ *does not correspond to any optimal solution of Model (OM). Thus, to capture all the supporting hyperplanes of* $\mathcal{T}_{VRS}$ *we need to consider both (IM) and (OM) models.*

Let $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o)$ be an optimal solution of Model (IM) or (OM) at a given efficient point $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. Since, by Theorem 1, $(\bar{\mathbf{u}}, -\bar{\mathbf{v}})$ specifies the normal vector of a supporting hyperplane of $\mathcal{T}_{\mathrm{VRS}}$ at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, it can be utilized to locally describe the technology transformation function as

---

[8] Charnes et al, (1986) grouped all the production points into six classes: E, E', F, NE, NE' and NF. Based on Definition 1 $E \cup E' \cup F$ contains efficient points, $E \cup E'$contains strong efficient points and $NE \cup NE' \cup NF$ contains interior points.

$$F(\mathbf{x},\mathbf{y}) \coloneqq \bar{\mathbf{u}}^T\mathbf{y} - \bar{\mathbf{v}}^T\mathbf{x} + \bar{u}_o, \tag{7}$$

where $F(\bar{\mathbf{x}},\bar{\mathbf{y}}) = 0$. Implementing the general rule for this specific linear transformation into (3), yields the following formula for SE in a DEA setting

$$\mathcal{E}(\bar{\mathbf{x}},\bar{\mathbf{y}}) = -\frac{\sum_{i=1}^{m}\frac{\partial F}{\partial x_i}x_i}{\sum_{r=1}^{s}\frac{\partial F}{\partial y_r}y_r}\Big|_{(\bar{\mathbf{x}},\bar{\mathbf{y}})} = \frac{\bar{\mathbf{v}}^T\bar{\mathbf{x}}}{\bar{\mathbf{u}}^T\bar{\mathbf{y}}}. \tag{8}$$

Since DEA offers a piecewise linear frontier, rather than an everywhere differentiable function, we may face multiple supporting hyperplanes at some frontier units. If the given point, $(\bar{\mathbf{x}},\bar{\mathbf{y}})$, is a relative interior of the intersection of a supporting hyperplane of $\mathcal{T}_{VRS}$ with the technology set, $\mathcal{T}_{VRS}, itself$. Then the normal vector of the supporting hyperplane defines a unique value for SE via (8). To accommodate the case of non-unique supporting hyperplanes Banker and Thrall (1992) proposed finding the upper and lower bounds for SE by applying models (IM) and (OM). For this, they fix the optimal objective of (IM) at unity by adding the constraint[9] $\mathbf{u}^T\bar{\mathbf{y}} + u_o = 1$ to (IM), and then account for both $\min\ u_o$ and $\max\ u_o$ on the optimal solutions of (IM). Let $u_{o\ IN}^-$ and $u_{o\ IN}^+$ be the minimum and maximum values of u_o, respectively. For measuring SE pertaining to every optimal solution of Model (IM), given that $\mathbf{v}^T\bar{\mathbf{x}} = 1,$ expression (8) can be simplified as

$$\mathcal{E}_{IN} = \frac{1}{1-\bar{u}_o}. \tag{9}$$

Now by applying (9) for $u_{o\ IN}^-$ and $u_{o\ IN}^+$, the input-oriented interval of SE is identified as $[\mathcal{E}_{IN}^-, \mathcal{E}_{IN}^+]$.

Implementing the same procedure for the output-oriented model (OM), it can be shown that for every optimal solution of (OM), (8) reduces to the following formula

$$\mathcal{E}_{OUT} = 1 + \bar{u}_o. \tag{10}$$

Similarly, one can compute the upper and lower values of u_o from all optimal solutions of Model (OM), say $u_{o\ OUT}^-$ and $u_{o\ OUT}^+$, and derive an output-oriented SE interval as $[\mathcal{E}_{OUT}^-, \mathcal{E}_{OUT}^+]$, by applying (10). The introduced input and output-oriented SE intervals may not be the same at some efficient points. Hence to consider all possible values for SE, the final SE interval will be $[\mathcal{E}^-, \mathcal{E}^+]$ by comparing both input and output SE intervals.[10] Note that an open interval will occur in the case of infinity.

2.2.2 Issues with existing methods

[9] This constraint ensures that the DMU under evaluation is located on the efficient frontier.

[10] See (Zelenyuk, 2013) for more information.

Figure 1 presents a simple example of a single input and single output technology for six DMUs.

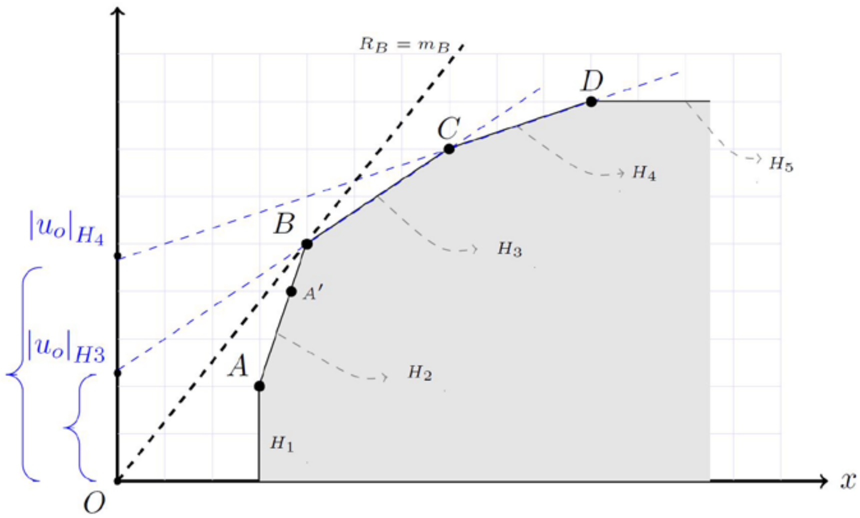

Figure 1: Efficiency frontier and hyperplanes

To provide more clarification, we report in the second column of Table 1, below, the average productivity (AP) of each DMU and in the third and fourth columns of Table 1 , respectively, all the *defining*[11] hyperplanes (*H1* to *H5)* of $\mathcal{T}_{\text{VRS}}$ passing through each DMU as well as the corresponding SEs calculated based on these hyperplanes.[12]

DMU	AP	Defining Hyp.	SE on Hyp.
$A = (3, 2)$	0.67	$H1: \; -x_1 + 3 = 0$ $H2: \; -0.75x + 0.25y + 1.75 = 0$	$\mathcal{E}_{A,H_1} = \mathcal{E}^+ = \infty$ $\mathcal{E}_{A,H_2} = \mathcal{E}^- = 4.5$
$A' = (3.67, 4)$	1.09	$H2: \; -0.75x + 0.25y + 1.75 = 0$	$\mathcal{E}_{A',H_2} = \mathcal{E} = 2.75$
$B = (4, 5)$	1.25	$H2: \; -0.75x + 0.25y + 1.75 = 0$ $H3: \; -0.40\ x + 0.60\ y - 1.4 = 0$	$\mathcal{E}_{B,H_2} = \mathcal{E}^+ = 2.4$ $\mathcal{E}_{B,H_3} = \mathcal{E}^- = 0.53$
$C = (7, 7)$	1.00	$H3: \; -0.40\ x + 0.60\ y - 1.4 = 0$ $H4: \; -0.25\ x + 0.75\ y - 3.5 = 0$	$\mathcal{E}_{C,H_3} = \mathcal{E}^+ = 0.67$ $\mathcal{E}_{C,H_4} = \mathcal{E}^- = 0.33$
$D = (10, 8)$	0.80	$H4: \; -0.25\ x + 0.75\ y - 3.5 = 0$ $H5: \; y - 8 = 0$	$\mathcal{E}_{E,H_4} = \mathcal{E}^+ = 0.41$ $\mathcal{E}_{E,H_5} = \mathcal{E}^- = 0$

Table 1: SE measures for DMUs shown in Figure 1.

First, consider unit A' in Fig. 1 where the only hyperplane supporting $\mathcal{T}_{\text{VRS}}$ at this point is H_2. The input and output multiplier values are $\text{v}_{H_2} = 0.75, u_{H_2} = 0.25$, respectively, hence

[11] A defining hyperplane of $\mathcal{T}_{\text{VRS}}$ at a given point is a supporting hyperplane such that its normal vector cannot be written as a convex combination of normal vectors of any other supporting hyperplanes of $\mathcal{T}_{\text{VRS}}$ at this point.

$$\mathcal{E}_{A',H_2,} = \frac{v_{H_2}x_{A'}}{u_{H_2}\times u_{A'}} = \frac{0.75\bullet 3.67}{0.2\bullet 5.4} = 2.75.$$

Also, applying each of (IM) and (OM) models along with (8) gives a unique SE value of 2.75 at A'. The problems arise from units that lie on multiple supporting hyperplanes, e.g., unit A in Fig. 1. There is an infinite number of supporting hyperplanes passing through A, where H_1 and H_2 specify the two defining ones. It can be shown that the normal vector of any supporting hyperplane passing through A is a convex combination of the H_1 and H_2 normal vectors. With relation to H_1, the SE value calculated from (8) becomes infinity. However, using H_2 the SE value is 4.5. Any supporting hyperplane between H_1 and H_2 gives an SE value greater than 4.5 until eventually the SE interval is determined as $[4.5, \infty)$.Thus, the first issue is whether 4.5 or any other value in the interval up to infinity should be selected as the SE for DMU A.

A second issue is that some solvers might return with status "unbounded" when maximizing $u_{o\ OUT}$ for DMU A, which is the case for GUROBI, or might return the direction in which the model is unbounded, i.e. $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o) = (0,1,3)$, which is the case for LINPROG. Consequently, this willincorrectly set a maximum SE value, $\mathcal{E}^+_{OUT} = 1 + 3 = 4$, noting that formula (10) only works if $\bar{\mathbf{u}}\bar{\mathbf{y}} = 1$ (while in this case $\bar{\mathbf{u}}\bar{\mathbf{y}} = 0$). It may also yield an inconsistency between the upper and lower bounds of SE, as in this case $u^-_{o\ OUT} = 4.5 > u^+_{o\ OUT}$. Similarly, for DMU D and $u^-_{o\ IN}$, GUROBI returns with status "unbounded" and LINPROG returns $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o) = (1,0,-8)$, which results in $\mathcal{E}^-_{IN} = \frac{1}{1-\bar{u}_o} = \frac{1}{1+8} = 0.11$. This too is an incorrect lower bound of SE since formula (9) only works if $\bar{\mathbf{v}}\bar{\mathbf{x}} = 1$.

Next, consider unit B for which the SE interval is [0.53,2.4]. The lower bound of SE, 0.53, indicates that the average productivity is greater than marginal productivity while the upper bound, 2.4, implies the opposite. Hence, choosing different values from the SE interval would result in varying and possibly contradicting types of RTS characterisation and SE assessment. For example, SE values of 2.4 and 0.53 indicate that B may be operating under either increasing or decreasing RTS, respectively. The method we develop in the next section removes this indeterminacy. We summarize the issues that may occur in measuring SE by conventional methods as follows:

(a) Typically, the transformation function of a DEA technology is piecewise linear where, at some strong efficient unit, there exist an infinite number of supporting hyperplanes, and thus existing methods identify an interval of possible SE values rather than a unique SE value. As such there are inherent ambiguities in choosing a single candidate from the aforementioned interval as the final SE value. This becomes an issue of paricular interest in characterising RTS when unity lies between the lower bound and upper bound of the SE interval, e.g. see unit B in Fig. 1 and its SE interval reported in Table 1.

(b) Choosing the lower bound as an indicator for SE, may yield the extreme value of zero, $\mathcal{E} = 0$, e.g. for a strong efficient DMU, D in Fig. 1. [13]

[13] In this case, some solvers return the status "unbounded" or an unrealistic value for the input oriented SE.

(c) Choosing the upper bound as an indicator for SE, may yield the extreme value of infinity, $\mathcal{E} = \infty$, e.g., unit A in Fig 1. [14]

The last two issues show that equiproportional changes in outputs corresponding to equiproportional changes in all inputs while moving toward the related hyperplane may be zero, or extend as high as infinity. For the strong efficient DMUs both results seem unreasonable from an economic perspective. For example, even a slight change in the inputs and outputs of unit A can lead to a massive shift of SE. This issue becomes even more important when we compare the SE value of the same units in different time periods.

(d) When dealing with input-output multipliers in Model (IM) or (OM), some solvers may fail -- as they did in our simulations -- to solve the models due to the large number of zero weights in the optimal solution The issue with zero weights is that they may lead to an unbounded optimal solution or unrealistic results. For example, it is often the case in practice that a DMU is unlikely to accept zero weights in inputs or outputs assigned by the model solution.

To overcome the issues mentioned above, we develop a unified approach for calculating SE, that offers a unique quantitative value for SE that can be used as a metric to compare different units, characterise the type of RTS and determine the MPSS units, simultaneously.

3 Most productive scale size and SE

3.1 Intuition for the new approach

The concept of MPSS introduced by (Banker, 1984) plays a pivotal role in our approach. Information about estimated MPSS allows decision makers to resolve questions about returns to scale for multiple inputs and output technologies without the need for additional information on prices or costs (see Banker et al., 2004). The MPSS for a given input and output mix is the scale size in which the outputs produced per unit of the inputs is maximised, (see Banker, 1984). In the other words, MPSS is a unit that maximises the average productivity for its given input-output vector.

Consider DMU B in Fig. 1 and Table 1. It can be seen that by slightly smoothing the frontier at B, the marginal productivity, m_B, exactly coincides with the average productivity, R_B in Fig. 1, thereby $\mathcal{E}_B = 1$. This DMU technically has the optimal scale known as the most productive scale size (MPSS), i.e., the average productivity ray, y/x, reaches its maximum value, and the production technology exhibits *constant returns to scale* (CRS). This is shown in Table 1 where the ratio y/x of DMU B, equals $5/4 = 1.25$, and is larger than that of the other DMUs. If SE is greater than unity, this means the equiproportional changes in outputs are greater than the equiproportional changes in inputs so in order to increase average productivity one would increase the scale size. On the other hand, if SE for a given DMU is less than unity then the proportional changes in its outputs is less than proportional changes in its inputs hence the

[14] In this case, some solvers return the status ``unbounded” or an unrealistic value for the output oriented SE.

average productivity can be increased with a smaller scale size. Thus, in both cases, in order to reach a desired scale, the units are getting closer to the MPSS part of the frontier.

Cooper et al. (2007) demonstrate that if $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ lies on the CRS part of the technology, then it is an MPSS and thus there is a supporting hyperplane passing through it in the form of $\bar{\mathbf{u}}^T\bar{\mathbf{y}} - \bar{\mathbf{v}}^T\bar{\mathbf{x}} = 0$ in which the value of $\bar{u}_o$ is zero. Equation (8) for this specific hyperplane implies that the SE value for any unit on this hyperplane is equal to one. Thus measuring SE with reference to the MPSS frontier reduces to finding a supporting hyperplane that offers an SE value close to unity, i.e., $\bar{\mathbf{u}}^T\bar{\mathbf{y}}$ is as close as possible to $\bar{\mathbf{v}}^T\bar{\mathbf{x}}$, or $|\bar{\mathbf{u}}^T\bar{\mathbf{y}} - \bar{\mathbf{v}}^T\bar{\mathbf{x}}|$ is as close as possible to zero. Equivalently, $|\bar{u}_o|$ is as close as possible to zero.

Therefore, moving from each efficient point towards the supporting hyperplane that has a smaller absolute value of u_o will lead to an improvement in average productivity, i.e., the hyperplane that has a smaller absolute value of $\bar{u}_o$, is closer to the MPSS frontier of the technology. For any given efficient unit, the $|\bar{u}_o|$ value of the supporting hyperplane can thus be viewed as a divergence indicator from the MPSS-frontier. This is defined using the notion of "closeness" to the MPSS frontier as follows:

Definition 2 *Suppose* $\bar{H} = \{(\boldsymbol{x}, \boldsymbol{y}): \bar{\boldsymbol{u}}^T\boldsymbol{y} - \bar{\boldsymbol{v}}^T\boldsymbol{x} + \bar{u}_o = 0\}$ *and* $\hat{H} = \{(\boldsymbol{x}, \boldsymbol{y}): \hat{\boldsymbol{u}}^T\boldsymbol{y} - \hat{\boldsymbol{v}}^T\boldsymbol{x} + \hat{u}_o = 0\}$ *are both supporting* $\mathcal{T}_{VRS}$ *at* $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$. *We say* $\bar{H}$ *is closer to the MPSS frontier than* $\hat{H}$ *if* $|\bar{u}_o| \leq |\hat{u}_o|$.

For example, consider DMU C in Fig. 1 which has supporting hyperplanes H_3 and H_4. In Table 1, $|u_o|$ of hyperplane H_3 is smaller than $|u_o|$ of hyperplane H_4 and of all other supporting hyperplanes of $\mathcal{T}_{\text{VRS}}$ passing through C. By Definition 2 Hyperplane H_3 is closer to the MPSS frontier than any other supporting hyperplane for DMU C. In order to increase average productivity, DMU C needs to decrease its scale size by moving toward H_3 that is closer to the MPSS, B, with $\mathcal{E}_C = 0.67$ rather than $\mathcal{E}_C = 0.33$.

For A and D, hyperplanes H_2 and H_4 respectively are closer to the MPSS frontier than any other supporting hyperplanes for these DMUs, offering a better improvement in the average productivity of A and D than any other hyperplanes passing through them. and the SE of A and D is thus measured as $\mathcal{E}_A = 4.5$ and $\mathcal{E}_D = 0.41$, respectively. Unit B is the MPSS and so $\mathcal{E}_A = 1$ with the ray, R_B, passing through this point.

3 Operationalisation

Let $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ be a given strong efficient point. The previous section established that SE can be measured at this point by identifying a supporting hyperplane that has a minimum distance from the MPSS frontier. To operationalize this idea, we introduce the following LP formulation:

$$
\begin{array}{lll}
\max & \rho & \\
\text{s.t.} & \mathbf{u}^T\mathbf{y}^j - \mathbf{v}^T\mathbf{x}^j + u_o \leq 0, & j = 1, \dots, n,, \\
& \mathbf{u}^T\bar{\mathbf{y}} - \mathbf{v}^T\bar{\mathbf{x}} + u_o = 0, & \\
& \rho \leq \mathbf{v}^T\bar{\mathbf{x}} \leq 1, & \\
& \rho \leq \mathbf{u}^T\bar{\mathbf{y}} \leq 1, & \\
& \mathbf{v}, \mathbf{u}, \rho \geq 0,\ u_o\ unrestricted. &
\end{array} \tag{11}
$$

The first set of $n-1$ inequality constraints followed by the equality constraint for the efficient point $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ ensures that this model obtains a supporting hyperplane of $\mathcal{T}_{\text{VRS}}$ at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. By choosing the optimal value of ρ, the model maximises the weighted inputs and outputs values simultaneously thereby locating the supporting hyperplanes for which $\mathbf{u}^T\bar{\mathbf{y}}$ is as close as possible to $\mathbf{v}^T\bar{\mathbf{x}}$; and thus gives the closest hyperplane to the MPSS frontier.

Let $(\mathbf{u}^*, \mathbf{v}^*, u_o^*, \rho^*)$ be an optimal solution of Model (11). By using (8), the value of SE obtained by Model (11) is

$$
\mathcal{E}_\rho = \frac{\mathbf{v}^{*T}\bar{\mathbf{x}}}{\mathbf{u}^{*T}\bar{\mathbf{y}}}. \tag{12}
$$

Corollary 2 below proves that this metric is *well-defined*.

Theorem 2 *At any strong efficient point of $\mathcal{T}_{VRS}$, Model (11) is feasible, bounded and has positive objective value, i.e., $\rho^* > 0$.*

Proof. See Appendix for the proof.

By Theorem 2, we can deduce that the value of SE calculated from (12), cannot be zero or infinity for a given efficient point.

Corollary 1 *For any strong efficient point of $\mathcal{T}_{VRS}$, $\mathcal{E}_\rho$ takes a finite positive value, i.e., $\mathcal{E}_\rho$ is neither zero nor infinity.*

Proof. Let $(\mathbf{u}^*, \mathbf{v}^*, u_o^*, \rho^*)$ be an optimal solution of Model (11). From Theorem 2 we have $\rho^* > 0$. This implies that both $\mathbf{v}^*\bar{\mathbf{x}}(\neq 0)$ and $\mathbf{u}^*\bar{\mathbf{y}}(\neq 0)$ have positive finite values. Thus their ratio has finite positive value and the proof is complete.

Lemma 1 shows that for any strong efficient unit of $\mathcal{T}_{\text{VRS}}$, $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, the optimal solution of Model (11) belongs to one of three different groups.

Lemma 1 *Let $(\boldsymbol{u}^*, \boldsymbol{v}^*, u_o^*, \rho^*)$ be an optimal solution of Model (11) for the strong efficient unit, $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$. Only one of the following conditions will hold:*

(a) $\rho^* = 1$ and $\mathbf{v}^{*T}\bar{\mathbf{x}} = \mathbf{u}^{*T}\bar{\mathbf{y}} = 1$.
(b) $\mathbf{v}^{*T}\bar{\mathbf{x}} = 1$ and $\mathbf{u}^{*T}\bar{\mathbf{y}} = \rho^* < 1$.

(c) $\mathbf{u}^{*T}\bar{\mathbf{y}} = 1$ and $\mathbf{v}^{*T}\bar{\mathbf{x}} = \rho^* < 1$.

Corollary 2 shows that the optimal value of Model (11) readily leads to defining $\mathcal{E}_\rho$ in (12). It also establishes that our proposed SE metric is "well-defined" in the sense that it is invariant to multiple optimal solutions of Model (11) because we are dealing with the optimal value of the LP model rather than its optimal solutions.

Corollary 2 *At any strong efficient unit of $\mathcal{T}_{VRS}$, $\mathcal{E}_\rho$ equals the optimal value of Model (11) or its reciprocal, i.e., $\mathcal{E}_\rho = \rho^*$ or $\mathcal{E}_\rho = \frac{1}{\rho^*}$.*

Proof. See Appendix for the proof.

Thus Lemma 1 and Corollary 2 show only one of the following results will hold in the optimal solution of Model (11) for any given strong efficient point:

(i) $\rho^* = 1, u_o^* = 0$ and $\mathcal{E}_\rho = 1$.
(ii) $\mathbf{u}^{*T}\bar{\mathbf{y}} = 1, \mathbf{v}^{*T}\bar{\mathbf{x}} = \rho^*, u_o^* < 0$ and $\mathcal{E}_\rho = \rho^* < 1$
(iii) $\mathbf{v}^{*T}\bar{\mathbf{x}} = 1, \mathbf{u}^{*T}\bar{\mathbf{y}} = \rho^*, u_o^* > 0$ and $\mathcal{E}_\rho = \frac{1}{\rho^*} > 1$.

To illustrate Corollary 1, Lemma 1 and Corollary 2, Table 2 reports the results of Model (11) using the data of Table 1. For DMU B, $\mathbf{v}^{*T}\bar{\mathbf{x}} = \mathbf{u}^{*T}\bar{\mathbf{y}} = 1 = \rho^* = \mathcal{E}_\rho$. For DMUs A and A', $\mathbf{v}^{*T}\bar{\mathbf{x}} = 1, \mathbf{u}^{*T}\bar{\mathbf{y}} = \rho^* < 1$ and $\mathcal{E}_\rho = \frac{1}{\rho^*} > 1$. For DMUs B and D, $\mathbf{u}^{*T}\bar{\mathbf{y}} = 1, \mathbf{v}^{*T}\bar{\mathbf{x}} = \rho^* < 1$ and $\mathcal{E}_\rho = \rho^* < 1$.

Results of Model (11) for data from Fig. 1

DMU	(x, y)	u^*	v^*	u_o^*	ρ	v^*x	u^*y	$1/\rho$	$\mathcal{E}_\rho$
A	$(3, 2)$	0.11	0.33	0.78	0.22	1.00	0.22	4.50	4.50
B	$(4, 5)$	0.20	0.25	0.00	1.00	1.00	1.00	1.00	1.00
C	$(7, 7)$	0.14	0.095	−0.33	0.67	0.67	1.00	1.50	0.67
D	$(10, 8)$	0.13	0.04	−0.58	0.42	0.42	1.00	2.40	0.42
A'	$(3.67, 4)$	0.09	0.27	0.64	0.36	1.00	0.36	2.75	2.75

We are now ready to prove our main theorem:

Theorem 3 *For any $(\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{u}_o)$ such that $\bar{\boldsymbol{u}}^T \boldsymbol{y}^j - \bar{\boldsymbol{v}}^T \boldsymbol{x}^j + \bar{u}_o \le 0$ for $j = 1, \dots, n$, $(\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}) \ge 0$ and $\bar{\boldsymbol{u}}^T \bar{\boldsymbol{y}} - \bar{\boldsymbol{v}}^T \bar{\boldsymbol{x}} + \bar{u}_o = 0$ we have $|u_o^*| \le |\bar{u}_o|$.*

Proof. See Appendix for the proof.

Theorem 3 proves that the supporting hyperplane at the given strong efficient DMU generated by the optimal solution of Model (11) is closer to the MPSS frontier than any other supporting hyperplanes of $\mathcal{T}_{\mathrm{VRS}}$ at this DMU.

3.2.1 Relationship with BCC models and RTS Classification

Let $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o)$ display the elements of a supporting hyperplane of $\mathcal{T}_{\mathrm{VRS}}$ at the given strong efficient point $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. If $u_o^* < 0$ $(u_o^* > 0)$, then $\bar{u}_o < 0$ $(\bar{u}_o > 0)$ for any corresponding supporting hyperplane of $\mathcal{T}_{\mathrm{VRS}}$ at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. Alternatively, if $u_o^* = 0$, then there is a supporting hyperplane of $\mathcal{T}_{\mathrm{VRS}}$ at this unit so that $\bar{u}_o = 0$. Theorem 4 formalises this relationship between u_o^* from Model (11) and $\bar{u}_o$ from Model (IM) (or (OM)).

Theorem 4 *Assume* $(\bar{\boldsymbol{u}}, \bar{\boldsymbol{v}}, \bar{u}_o)$ *is an optimal solution of Model (IM). Then we have*

(i) $u_o^* < 0$ if and only if $\bar{u}_o < 0$ for all optimal solutions of Model (IM);
(ii) $u_o^* > 0$ if and only if $\bar{u}_o > 0$ for all optimal solutions of Model (IM);
(iii) $u_o^* = 0$ if and only if $\bar{u}_o = 0$ for some optimal solution of Model (IM).

A similar result holds for Model (OM).

Proof. See Appendix for the proof.

Next we use Theorem 5 to establish the optimal solution of Model (11) can reliably identify RTS.

Theorem 5 *For any strong efficient unit,* $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$*, one of the following is satisfied;*

(i) IRS prevails at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ if and only if $\mathcal{E}_\rho > 1$.
(ii) DRS prevails at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ if and only if $\mathcal{E}_\rho < 1$.
(iii) CRS prevails at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ if and only if $\mathcal{E}_\rho = 1$.

Proof. See Appendix for the proof.

3.3 Properties and Advantages

The properties and advantages of Model (11) are as follows.

P1. Non-oriented and frontier based

Scale elasticity is a characteristic of the frontier of the technology. Even studies that investigate SE for the interior points of the technology set calculate SE with the respective projection point located on the frontier. Our approach is frontier based and thus does not depend on the orientation whether input or output or directional. This aligns with Zelenyuk (2013) who demonstrates that SE measures based on the directional distance function are equivalent to the input oriented and output oriented SE measures for the efficient points.

P2. MPSS based SE

Our model operationalises the concept of *closeness* to the MPSS frontier for the measurement of SE .

P3. Unified criterion

Our single LP model specifies a single numerical value that is economically meaningful for scale elasticity as opposed to the intervals used in other methods and associated indeterminacy problems. The single numerical value, ρ, is either directly or inversely related to SE. This unified criterion is invariant to changes in input and output values and is consistent across LP solvers because we are dealing with the optimal value of an LP model, which is always greater than or equal to zero, rather than the optimal solution of the model.

P4. RTS type identification

The optimal value of our proposed model can classify the type of RTS for each unit without any additional calculations or use of additional models. Furthermore, it does so correctly in the sense of Definition 2.

P5. DMU classification

Section 2.2.1 provides theorems and results only for strong efficient points. However our proposed model can be used without prior knowledge of efficient and inefficient DMUs, i.e., it obtains a positive SE value for efficient DMUs and zero for inefficient DMUs as per Remark 2. In other words, Model (11) classifies DMUs into strong efficient, efficient and inefficient points.

Remark 2 *At any point of $\mathcal{T}_{VRS}$ that is not a strong efficient point, Model (11) is feasible, bounded and has the value zero as its objective value, i.e., $\rho^* = 0$.* More precisely, assume $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a point of $\mathcal{T}_{\text{VRS}}$ and Model (11) has an optimal solution $(\mathbf{u}^*, \mathbf{v}^*, u_o^*, \rho^*)$. Then one of the following will hold:

(i) $\rho^* > 0$ if and only if $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a strong efficient point of $\mathcal{T}_{\text{VRS}}$.
(ii) $\rho^* = 0$ and $(\mathbf{u}^*, \mathbf{v}^*) \neq \mathbf{0}$ if and only if $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is an efficient point of $\mathcal{T}_{\text{VRS}}$ but it is not a strong efficient point.
(iii) $\rho^* = 0$ and $(\mathbf{u}^*, \mathbf{v}^*, u_o^*) = \mathbf{0}$ if and only if $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is an interior point of $\mathcal{T}_{\text{VRS}}$.[15]

4 Illustrated Example

[15] Note that the result, viz. $\rho^* = 0$, in (ii) and (iii) above is not restrictive since all DMUs can be projected to a strong efficient point along the frontier before SE can be calculated.

In this section we present numerical results obtained by our proposed Model (11) and compare these with the standard input and output oriented models under VRS. [16] We select 95 Commercial Banks with four variables as inputs and three as outputs.[17] First we consider the technology set, $\mathcal{T}_{\mathrm{VRS}}$, in each year from 2001 to 2011 constructed by the 95 banks in the sample. We then solve our model for all the banks but demonstrate the results only for the strong efficient units across the 11 years. In this example, there are 10 DMUs (banks) on the strong efficient frontier of $\mathcal{T}_{\mathrm{VRS}}$ in years 2001 to 2011. These results are shown in Table 2. Superscripts I and D in Table 1 stand for IRS and DRS, receptively, and values of 1.00 indicate MPSS.

DMU	2001	**2002**	2003	**2004**	2005	**2006**	2007	**2008**	2009	**2010**	2011
12	1.44^{I}	1.34^{I}	1.17^{I}	1.11^{I}	1.84^{I}	1.20^{I}	1.93^{I}	2.32^{I}	1.65^{I}	1.63^{I}	1.52^{I}
14	1.24^{I}	1.12^{I}	1.06^{I}	1.03^{I}	1.11^{I}	1.20^{I}	1.19^{I}	1.24^{I}	1.09^{I}	1.16^{I}	1.01^{I}
29	1.03^{I}	1.03^{I}	1.06^{I}	1.05^{I}	1.02^{I}	1.13^{I}	1.39^{I}	1.57^{I}	1.21^{I}	1.25^{I}	1.22^{I}
37	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99^{D}	1.00
43	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
62	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
69	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
79	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
89	1.00	1.00	1.00	1.00	1.00	1.00	0.98^{D}	0.97^{D}	0.99^{D}	0.99^{D}	0.99^{D}
91	2.39^{I}	1.93^{I}	1.38^{I}	1.00	1.19^{I}	1.00	1.00	1.00	1.00	1.00	1.00

Table 2: SE results obtained from Model (11) for strong efficient banks in 2001-2011

To illustrate the issues outlined in Section 2.2.1, we have computed the input and output-oriented SE intervals for DMUs 12 and 37. They are obtained by formulas (9), (10) and the related LP models are solved by two commonly used solvers, GUROBI and LINPROG.

		2009		2010		2011	
Solvers	DMU	$[\mathcal{E}_{IN}^{-},\ \mathcal{E}_{IN}^{+}]$	$[\mathcal{E}_{OUT}^{-},\ \mathcal{E}_{OUT}^{+}]$	$[\mathcal{E}_{IN}^{-},\ \mathcal{E}_{IN}^{+}]$	$[\mathcal{E}_{OUT}^{-},\ \mathcal{E}_{OUT}^{+}]$	$[\mathcal{E}_{IN}^{-},\ \mathcal{E}_{IN}^{+}]$	$[\mathcal{E}_{OUT}^{-},\ \mathcal{E}_{OUT}^{+}]$
(GUROBI)	12	$[1.65, \infty)$	$[1.65, \infty)$	$[1.63, \infty)$	$[1.63, \infty)$	$[1.52, \infty)$	$[1.52, \infty)$
	37	$(-\infty, 1.00]$	$[0.00, 1.00]$	$(-\infty, 0.99]$	$[0.00, 0.99]$	$(-\infty, 1.00]$	$[0.00, 1.00]$
(LINPROG)	12	$[1.65, \infty)$	$[1.65, \mathbf{1.38}^{*}]$	$[1.63, \infty)$	$[1.63, \mathbf{1.39}^{*}]$	$[1.52, \infty)$	$[1.52, \mathbf{1.36}^{*}]$
	37	$[\mathbf{0.22}^{*}, 1.00]$	$[0.00, 1.00]$	$[\mathbf{0.23}^{*}, 0.99]$	$[0.00, 0.99]$	$[\mathbf{0.22}^{*}, 1.00]$	$[0.00, 1.00]$

Table 3: Input and output-oriented SE intervals for strong efficient banks in 2009-2011

[16] This data set is extracted from Datastream and it is a balanced panel of large European banks for the period 2001 to 2011.

[17] The inputs are fixed assets, personnel expenses, total customer deposits, and equity. The outputs are loans, other earning assets and off balance sheet assets.

It can be seen from Table 3 that different solvers can sometimes produce different results for the same banks. For example, comparing DMU 12 2009 results from the two different solvers, reveals that GUROBI returns with status “unbounded” whereas LINPROG yields inconsistent SE values.[18] Inconsistent results are highlighted in bold and with an asterisk in Table 3.

Table 3 clearly illustrates the problems arising with conventional methods as discussed in Section 2.2.2, in particular, issues (b), (c) and (d). In comparison, our proposed model produces consistent results across solvers, as shown in Table 2.

Summary

Scale elasticity (SE) is an important feature of production technology and plays an important role in decision making, e.g. mergers and acquisitions, and in shaping competitive market structure and market conduct. Hence it is important that it is measured corectly. While there have been several attempts to develop useful measures of SE, due to the linear piecewise nature of the DEA technology these have resulted in interval measures with upper and lower bounds . Such intervals have led to other issues of indeterminacy and computational limits and inconsistencies.

This paper has proposed an LP model providing a unique measure of SE that is based on a simple proposition of closeness to MPSS. The model is non-oriented and the result not only provides the SE measure but also allows returns to scale to be classified. Furthermore, the model provides feasible consistent results using different solvers, while the same can not be said for the results of other approaches. The model was tested using a data set of 95 banks across 11 years and compared with the interval approaches which provided conflicting and inconsistent results.

Appendix

Proof of Theorem *2*

Proof. Suppose $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is a given strong efficient point of $\mathcal{T}_{VRS}$. By Definition 1 there exists a vector $(\bar{u},\bar{v}) > 0$ along with a scalar $\bar{u}_o$ that satisfies (5). Hence $\bar{\mathbf{v}}\bar{\mathbf{x}} \neq 0$ and $\bar{\mathbf{u}}\bar{\mathbf{y}} \neq 0$. Dividing $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o)$ by $\bar{\mathbf{v}}\bar{\mathbf{x}} + \bar{\mathbf{u}}\bar{\mathbf{y}}$, with $\bar{\mathbf{v}}\bar{\mathbf{x}} > 1$ or $\bar{\mathbf{u}}\bar{\mathbf{y}} > 1$, the new vector $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{u}_o, \hat{\rho})$ constructs a feasible solution for Model (11) where $\hat{\rho} = \min\{\hat{\mathbf{v}}\bar{\mathbf{x}}, \hat{\mathbf{u}}\bar{\mathbf{y}}\} > 0$. Hence Model (11) is feasible and its optimal value is positive. The third (and fourth) set of constraints in (11) guarantees the boundedness of this model since ρ is limited by one.

Proof of Lemma *1*

[18] Note the upper bound is larger than the lower bound in the estimated SE interval non-withstanding that the obtained value is not calculated correctly since (9) holds only if $\mathbf{v}^T\bar{\mathbf{x}} = 1$ which is not the case here..

Proof. Since Model (11) maximises ρ, there are two cases:

(i) $\mathbf{u}^{*T}\bar{\mathbf{y}} = \rho^*$; in this case, if $\mathbf{v}^{*T}\bar{\mathbf{x}} = \rho^*$ then we conclude $\rho^* = 1$. Hence (C1) holds for $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. Suppose $\mathbf{v}^{*T}\bar{\mathbf{x}} > \rho^*$. Then by assumption we have $0 < \mathbf{u}^{*T}\bar{\mathbf{y}} < \mathbf{v}^{*T}\bar{\mathbf{x}}$. We prove $\mathbf{v}^{*T}\bar{\mathbf{x}} = 1$, and hence (C2) holds for $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ which is a desired result. By contradiction assume that $\mathbf{v}^{*T}\bar{\mathbf{x}} < 1$; dividing $(\mathbf{u}^*, \mathbf{v}^*, u_o^*)$ by $\mathbf{v}^{*T}\bar{\mathbf{x}}$ yields a new vector $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o)$ satisfying the first set of n-1 inequalities and the adjacent equation in (11) . Now let $\bar{\rho} = \bar{\mathbf{u}}^T\bar{\mathbf{y}}$. It follows that $(\mathbf{u}^*, \mathbf{v}^*, u_o^*, \bar{\rho})$ constructs a feasible solution for Model (11) since $\bar{\rho} < \bar{\mathbf{v}}^T\bar{\mathbf{x}} = 1$ and $\bar{\rho} = \bar{\mathbf{u}}^T\bar{\mathbf{y}} < 1$. Then we have $\bar{\rho} = \bar{\mathbf{u}}^T\bar{\mathbf{y}} = \frac{\mathbf{u}^{*T}\bar{\mathbf{y}}}{\mathbf{v}^{*T}\bar{\mathbf{x}}} > \mathbf{u}^{*T}\bar{\mathbf{y}} = \rho^*$ and thus $\rho^* < \bar{\rho}$, which contradicts the optimality of ρ^*.

(ii) $\mathbf{v}^{*T}\bar{\mathbf{x}} = \rho^*$; The proof is similar to (i).

Proof of Corollary 2

Proof. By Lemma 1 we have three cases to consider. (C1) $\mathbf{v}^{*T}\bar{\mathbf{x}} = \mathbf{u}^{*T}\bar{\mathbf{y}} = 1$ follows that $\mathcal{E}_\rho = \frac{\mathbf{v}^{*T}\bar{\mathbf{x}}}{\mathbf{u}^{*T}\bar{\mathbf{y}}} = 1$; (C2) $\mathbf{v}^{*T}\bar{\mathbf{x}} = 1$ and $\mathbf{u}^{*T}\bar{\mathbf{y}} = \rho^* < 1$ follows that $\mathcal{E}_\rho = \frac{\mathbf{v}^{*T}\bar{\mathbf{x}}}{\mathbf{u}^{*T}\bar{\mathbf{y}}} = \frac{1}{\rho^*} > 1$; (C3) $\mathbf{u}^{*T}\bar{\mathbf{y}} = 1$ and $\mathbf{v}^{*T}\bar{\mathbf{x}} = \rho^* < 1$ follows that $\mathcal{E}_\rho = \frac{\mathbf{v}^{*T}\bar{\mathbf{x}}}{\mathbf{u}^{*T}\bar{\mathbf{y}}} = \rho^* < 1$.

Proof of Theorem 3

Proof. By Lemma 1 we have three cases to consider. First, suppose $\mathbf{v}^{*T}\bar{\mathbf{x}} = \rho^* < 1$ and $\mathbf{u}^{*T}\bar{\mathbf{y}} = 1$ which implies $u_o^* < 0$. Consequently, we have $\bar{u}_o < 0$ for any optimal solution of (IM). To prove this, by contradiction, suppose $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o)$ is a solution for (IM) and that $\bar{u}_o > 0$. Since $\bar{\mathbf{v}}^T\bar{\mathbf{x}} = 1$ and $\bar{\mathbf{u}}^T\bar{\mathbf{y}} < 1$ dividing $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o)$ by $\bar{\mathbf{u}}^T\bar{\mathbf{y}}$ we normalize it to a feasible solution for Model (11) as $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{u}_o, \hat{\rho})$ where $\hat{\rho} = \hat{\mathbf{v}}^T\bar{\mathbf{x}} > 1$. This contradicts with the optimality of ρ^*. Hence $\bar{u}_o < 0$ for any optimal solution of (IM). Now, we prove $|u_o^*| \leq |\bar{u}_o|$. By contradiction suppose $|u_o^*| > |\bar{u}_o|$. Again dividing $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o)$ by $\bar{\mathbf{u}}^T\bar{\mathbf{y}}$ and letting $\hat{\rho} = \hat{\mathbf{v}}^T\bar{\mathbf{x}}$ we construct $(\hat{\mathbf{u}}, \hat{\mathbf{v}}, \hat{u}_o, \hat{\rho})$ as a feasible solution for Model (11). Since $\bar{\mathbf{u}}^T\bar{\mathbf{y}} > 1$, and from our assumption, we have $|u_o^*| \geq |\bar{u}_o| \geq |\hat{u}_o|$. It implies $u_o^* < \hat{u}_o$ and then $\mathbf{v}^{*T}\bar{\mathbf{x}} - \mathbf{u}^{*T}\bar{\mathbf{y}} < \hat{\mathbf{v}}^T\bar{\mathbf{x}} - \hat{\mathbf{u}}^T\bar{\mathbf{y}}$. Now substituting $\mathbf{u}^{*T}\bar{\mathbf{y}}$ and $\hat{\mathbf{u}}^T\bar{\mathbf{y}}$ with 1 follows $\rho^* < \hat{\rho}$. This result contradicts the optimality of ρ^*. Hence for this part of the frontier we have $|u_o^*| \leq |\bar{u}_o|$.

Now suppose $\mathbf{v}^{*T}\bar{\mathbf{x}} = 1$ and $\mathbf{u}^{*T}\bar{\mathbf{y}} = \rho^*$ which implies $u_o^* > 0$. Similar to the first case, it can be proven that $\bar{u}_o > 0$ for any optimal solution of (IM), which follows that $|u_o^*| \leq |\bar{u}_o|$.

In the case of $\mathbf{v}^{*T}\bar{\mathbf{x}} = \mathbf{u}^{*T}\bar{\mathbf{y}} = 1$ the proof is straightforward.

Note that we can consider the output oriented BCC model and prove the theorem in a similar way.

Proof of Theorem 4

Proof.

(i). In the proof of Theorem 3 we state $\bar{u}_o < 0$ for any optimal solution of (IM) if $u_o^* < 0$. Hence to complete the proof, assume $\bar{u}_o < 0$ for any optimal solution of (IM), we then need to prove $u_o^* < 0$. We can prove this by contradiction. First, suppose $u_o^* > 0$ that confirms $\bar{u}_o > 0$ for any optimal solution of (IM), which is contradicts our assumption. On the other hand, $u_o^* = 0$ implies that $(\mathbf{u}^*, \mathbf{v}^*, 0)/({\mathbf{v}^*}^T\bar{\mathbf{x}})$ is an optimal solution of (IM) which again contradicts the assumption.

(ii) The proof is similar to (i).

(iii) Let $u_o^* = 0$. The proof is straightforward since $(\mathbf{u}^*, \mathbf{v}^*, u_o^*)/({\mathbf{v}^*}^T\bar{\mathbf{x}})$ is an optimal solution of (IM). Now, suppose there is an optimal solution for (IM) as $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o)$ such that $\bar{u}_o = 0$. Hence we have $\bar{\mathbf{u}}^T\bar{\mathbf{y}} - \bar{\mathbf{v}}^T\bar{\mathbf{x}} = 0$, which follows that $(\bar{\mathbf{u}}, \bar{\mathbf{v}}, \bar{u}_o, 1)$ is a feasible solution for Model (11). This result contradicts $\rho^* < 1$.

Proof of Theorem 5

Proof. Banker and Thrall (1992), in Proposition 3, illustrate how $\bar{u}_o$ in the optimal solutions of Model (IM) or (OM), characterise the RTS of a given unit. We reformulate their proposition as follows:

Considering Model (IM) with a strong efficient point, $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$:

(I) IRS prevails at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ if and only if $\bar{u}_o > 0$ for all optimal solutions.
(II) DRS prevails at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ if and only if $\bar{u}_o < 0$ for all optimal solutions.
(III) CRS prevails at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ if and only if $\bar{u}_o = 0$ for some optimal solution.

Naturally, this is also true for Model (OM). We are now ready to prove Theorem 5.

(i) Suppose IRS prevails at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, then by taking account of Item (I), $\bar{u}_o > 0$ for all optimal solutions of the BCC model and, by Theorem 4, we have $u_o^* > 0$. Consequently ${\mathbf{v}^*}^T\bar{\mathbf{x}} > {\mathbf{u}^*}^T\bar{\mathbf{y}}$ since ${\mathbf{u}^*}^T\bar{\mathbf{y}} - {\mathbf{v}^*}^T\bar{\mathbf{x}} + u_o^* = 0$ which implies (c) in Lemma 1 holds for this unit and so ${\mathbf{v}^*}^T\bar{\mathbf{x}} = \rho^*$, ${\mathbf{u}^*}^T\bar{\mathbf{y}} = 1$ and $\mathcal{E}_\rho > 1$.

Now suppose $\mathcal{E}_\rho > 1$, which implies ${\mathbf{u}^*}^T\bar{\mathbf{y}} = 1$ and ${\mathbf{v}^*}^T\bar{\mathbf{x}} = \rho^*$. Hence, $u_o^* > 0$ and then by Theorem 4 we have $\bar{u}_o > 0$ for all optimal solution of (IM). This result indicates IRS for $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, by Item (I), and the proof is complete.

(ii) The proof for (ii) is similar to (i).

(iii) Suppose CRS prevails at $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, which implies that $u_o^* = 0$ by Theorem 4 and

Item (III). Then Lemma 1 indicates $\mathcal{E}_\rho = 1$.

Now suppose $\mathcal{E}_\rho = 1$ which implies $u_o^* = 0$, by Lemma 1. Then from Theorem 4 follows that $\bar{u}_o = 0$ for an optimal solution of (IM). This result, by taking account of Item (III), indicates CRS for $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$.

References

Balk, B. M., Färe, R., & Karagiannis, G. (2015). On directional scale elasticities. *Journal of Productivity Analysis*, 43, 99-104.

Banker, R. D. (1980). A game theoretic approach to measuring efficiency. European *Journal of Operational Research*, 5 (4), 262-266.

Banker, R. D. (1984). Estimating most productive scale size using data envelopment analysis. *European Journal of Operational Research*, 17 (1), 35-44.

Banker, R. D., Charnes, A., & Cooper, W. W.(1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30 (9), 1078-1092.

Banker, R. D., Cooper, W. W., Seiford, L. M., Thrall, R. M., & Zhu, J. (2004). Returns to scale in different DEA models. *European Journal of Operational Research*, 154(2), 345-362.

Banker, R. D., & Thrall, R. M. (1992). Estimation of returns to scale using data envelopment analysis. *European Journal of Operational Research*, 62 (1), 74-84.

Briec, W., & Leleu, H. (2003). Dual Representations of Non-Parametric Technologies and Measurement of Technical Efficiency. *Journal of Productivity Analysis*, 20 (1), 71-96.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data Envelopment Analysis*. Boston, MA: Springer US.

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1985). *The measurement of efficiency of production*. Boston: Kluwer-Nijhoff.

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1988). Scale Elasticity and Scale Efficiency. *Journal of Institutional and Theoretical Economics* (JITE), 144 (4), 721-729.

Førsund, F. R., & Hjalmarsson, L. (2004). Calculating scale elasticity in DEA models. *Journal of the Operational Research Society*, 55 (10), 1023-1038.

Fukuyama, H. (2000). Returns to scale and scale elasticity in data envelopment analysis. *European Journal of Operational Research*, 125 (1), 93-112.

Fukuyama, H. (2003). Scale characterizations in a DEA directional technology distance function framework. *European Journal of Operational Research*, 144 (1), 108-127.

Golany, B., & Yu, G. (1997. Estimating returns to scale in DEA. *European Journal of*

Operational Research, 103 (1), 28-37.

Panzar, J. C., & Willig, R. D. (1977). Economies of Scale in Multi-Output Production. *The Quarterly Journal of Economics*, 91 (3), 481.

Podinovski, V. V., F˙rsund, F. R., & Krivonozhko, V. E. (2009). A simple derivation of scale elasticity in data envelopment analysis. *European Journal of Operational Research*, 197 (1), 149-153.

Shepherd, R. W. (1970). *Theory of cost and production functions*. Princeton: Princeton University Press.

Starrett, D.A. (1977). Measuring returns to scale in the aggregate, and the scale effect of public goods. *Econometrica* 45, 1439–1455.

Sueyoshi, T. (1999). DEA Duality on Returns to Scale (RTS) in Production and Cost Analyses: An Occurrence of Multiple Solutions and Differences between Production-Based and Cost-Based RTS Estimates. *Management Science*, 45 (11), 1593-1608.

Zelenyuk, V. (2013). A scale elasticity measure for directional distance function and its dual: Theory and DEA estimation. *European Journal of Operational Research*, 228 (3), 592-600.