MEMORANDUM

No 10/2015

Economic Perspectives on DEA

Finn Førsund

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/econ
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 09/15	Andreas Kotsadam, Eivind Hammersmark Olsen, Carl Henrik Knutsen and Tore Wig *Mining and Local Corruption in Africa*
No 08/15	Eric Nævdal *Catastrophes and Expected Marginal Utility – How the Value of The Last Fish in a Lak is Infinity and Why We Shouldn't Care (Much)*
No 07/15	Niklas Jakobsson and Andreas Kotsadam *The Economics of Trafficking for Sexual Exploitation*
No 06/15	Geir B. Asheim and Stéphane Zuber *Evaluating Intergenerational Risks: Probability Adjusted Rank-Discounted Utilitarianism*
No 05/15	Fridrik Mar Baldursson and Nils-Henrik von der Fehr *Natural Resources and Sovereign Expropriation*
No 04/15	Erik Biørn and Xuehui Han *Persistence, Signal-Noise Pattern and Heterogeneity in Panel Data: With an Application to the Impact of Foreign Direct Investment in GDP*
No 03/15	Alice Ciccone *Environmental Effects of a Vehicle Tax Reform: Empirical Evidence from Norway*
No 02/15	Katinka Holtsmark and Kristoffer Midttømme *The Dynamics of Linking Permit Markets*
No 01/15	Francesco Lania and Alessia Russo *Public Education and Pensions in Democracy: A Political Economy Theory*
No 29/14	Lars Kirkebøen, Edwin Leuven and Magne Mogstad *Field of Study, Earnings, and Self-Selection*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/english/research/unpublished-works/working-papers/

Economic perspectives on DEA*

by

Finn R Førsund

Department of Economics, University of Oslo,

and

Norwegian Defence Research Establishment (FFI)

Memo 10/2015-v1

This version April 2015

Abstract: Research on productive efficiency at the firm level has developed as an important and active strand of research the last decades, both within operations research, management science and economics. However, the interests pursued within the fields have had some differences regarding sound theoretical foundations. The perspective within economics is highlighted and a critique of some research directions judged as unfortunate is offered.

Keywords: Efficiency measures; data envelopment analysis; shadow prices; weight restrictions; cross-efficiency

JEL-classification: C61, D24

* The paper is written within the research programme "Kostnadseffektiv drift av Forsvaret (KOSTER IV)" (Cost Efficiency in Defence) at the Norwegian Defence Research Establishment (FFI). I am indebted to Torbjørn Hanson and Kenneth Løvold Rødseth for valuable comments.

1. Introduction

Measuring productive efficiency has been developing during the last decades to become an important research strand within the fields of economics, management science and operations research. Two seminal contributions are Farrell (1957) and Charnes et al. (1978). Although the latter paper uses the efficiency definition of the former paper the approaches for setting up the calculation of the measures differ in the two papers and this has had consequences for how the research strand has developed within economics, management science and operations research due to the fact that researchers with the latter background have tended to neglect the contribution of Farrell (Førsund and Sarafoglou, 2002).

The purpose of the paper is to explore some basic issues in the efficiency literature with a focus on differences between economic interpretations and more operations-research based formulations of efficiency. The paper may therefore be especially useful for researchers from other disciplines than economics. The paper may also serve as an introduction to Data Envelopment Analysis (DEA) (the name was coined in Charnes et al., 1978), elucidating pitfalls and giving the basic ideas of efficiency analyses using DEA as a tool for researchers not so familiar with efficiency analysis and DEA.

The plan of the paper is as follows. In Section 2 the definitions of efficiency measures in the two seminal papers Farrell (1957) and Charnes et al. (1978) are reviewed and the differences in approaches pointed out. Section 3 explores the interpretations of the shadow prices (dual variables or weights) in DEA problems set up as linear programming problems. Section 4 provides general productivity interpretations of efficiency measures. Sections 5 and 6 take a look at developments within the DEA literature that may be of doubtful value: Section 5 reviews the introduction of restrictions on weights and Section 6 discusses the soundness of the cross-efficiency approach. Conclusions are offered at the end of each section and a general summing-up is presented in Section 7.

2. Efficiency measure definitions

Introduction

The fundamental definition of efficiency is apparently different in the two seminal papers; Farrell (1957) basing his definition on the proportional scaling needed for observations of an inefficient units to be projected onto an efficient production function (later termed the frontier function) and Charnes et al. (1978) basing their definition on an index of weighted outputs on weighted inputs, restricting this ratio to be less than or equal to the one for the most efficient operation (normalised to 1).

In order to carry out a study of efficiency three elements must be in place

i) Definitions of efficiency measures
ii) Methods for calculating efficiency measures
iii) Relevant data for inputs and outputs of an activity we want to measure efficiency for.

Farrell (1957) and Charnes et al. (1978) provided both definitions and methods. However, when Farrell followed up a comment when his paper was discussed at the presentation in Royal Statistical Association in 1976 that the newly developed linear programming could be applied, the application in Farrell and Fieldhouse (1962) was designed only to accommodate the case of a single output and constant returns to scale. Charnes et al. (1978) generalised the linear programming to multiple outputs and stressed the connection between dual and primal problems. These formulations are still in use today and may explain why this paper only is followed by researcher with management science and operational research background.

The need for relevant data was stressed in both seminal papers. A model representation of the real world must always be a simplification. There is a trade-off between focusing on a limited set of variables and the degrees of freedom from a methodological point of view. Suffice it to say that a fruitful choice of variables depends on knowledge of the activities at hand, as do the ability to draw conclusion from the analysis being of value for practical policy measures. The question of having sufficient data will be touched upon in Section 5.

One should be aware of the fact that it has been very difficult to establish the causes of efficiency differences between the units of analysis both from a theoretical and methodological point of view[1]. Without this knowledge it is very difficult to give advice on policies to improve efficiency. However, this issue is outside the scope of the paper (see e.g. Førsund (2010) for a discussion).

The Farrell definitions of efficiency measures

Farrell (1957) did not present formal definitions of the efficiency measures, maybe due to a wish to keep the exposition simple in order to be "of interest to a wide range of economic statisticians, business men and civil servants, many of whom have little knowledge of economic theory or mathematics" (p. 11). However, his widely reproduced graphical illustrations convey very well his efficiency definitions. His original illustration; Diagram 1, is reproduced in Figure 1. The variables *y* and *x* along the axes are input coefficients, i.e. an

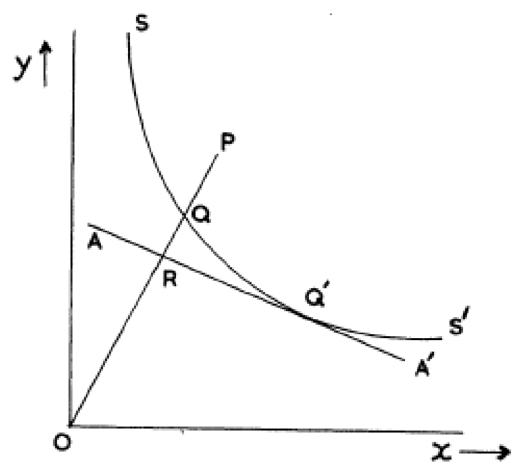

DIAGRAM 1.

Figure 1. The original Farrell (1957) illustration of efficiency measures

input divided by the output. The illustration is based on a single output and two inputs. The curve SS' represents the unit isoquant of the efficient, or frontier, production function assumed to have constant returns to scale (CRS)[2]. All technical information about the production function then collapses to this unit isoquant in input coefficient space. The unit under study is located at P. Technical efficiency is then defined as the relative distance to the

[1] Using two-stage procedures to seek for variables correlated with efficiency scores is more explorative than based on theory and a key issue is whether such environmental variables influence only the efficiency score or also the production possibility set.

[2] CRS means that expanding the inputs proportionally with the same factor the output(s) will increase with the same proportion.

frontier keeping output constant but shrinking the use of inputs proportionally in order to reach point Q on the frontier; OQ/OP.

Knowledge about a frontier technology can be obtained in two ways; based on engineering information about the production activity in question, or based on observed best practice. The latter approach is usually followed in empirical applications.

The Farrell input- and output-oriented efficiency measures can, in the case of multiple outputs and inputs, be generalised using a standard transformation function between inputs and outputs in economics, $F(Y,X)\,(F'_Y > 0, F'_X < 0)$, where X is a vector of inputs and Y a vector of outputs. We can use the expression $F(Y,X) \le 0$ as the representation of the general production possibilities expressing the set in general as

$$T = \{(Y,X) : Y \geq 0 \text{ can be produced by } X \geq 0\} \tag{1a}$$

The border of the set corresponds to a frontier production function. A formulation using the transformation function above is

$$T = \left\{(Y,X) : F(Y,X) \le 0, Y \ge 0, X \ge 0\right\} \tag{1b}$$

The last equation is derived inserting the transformation relation above. Equality in the transformation relation means that we are on the border of the set, i.e. on the frontier function, while strict inequality means we have an inefficient unit. For a unit j the formal definition of technical efficiency is:

$$E_{1j} = \text{Min}_{\theta_j}\left\{\theta_j : F(Y_j, \theta_j X_j) = 0\right\}, E_{1j} \in (0,1], j = 1,...,n \tag{2}$$

Farrell used the notation E_1 for this input-oriented technical efficiency measure.[3] A common frontier technology for all units is assumed.

Farrell also discussed an output-oriented technical efficiency measure defined as a scaling of outputs keeping inputs constant. He also wanted this measure E_2 to be between 0 and 1:

$$E_{2j} = Min_{\varphi_j}\left\{\varphi_j : F(\frac{1}{\varphi_j}Y_j, X_j) = 0\right\}, E_{2j} \in (0,1], j = 1,...,n \tag{3}$$

A separate diagram was not offered for the output-oriented measure, but that is not necessary in the CRS case because as Farrell pointed out the measures are identical in this case. (Point P is moved to the frontier point Q on SS' in Diagram 1 increasing the output for given inputs, with SS' now representing the isoquant for the output obtainable on the frontier.)

[3] Farrell (1957, p. 259) used the lowercase notation e_1 and e_2, but changed to upper-case notation for the input-oriented efficiency measure in Farrell and Fieldhouse (1962, p. 258).

The Farrell efficiency measures correspond to the concept of distance functions introduced in Shephard (1970). (Shephard's input distance function is the inverse of Farrell's input-oriented efficiency measure, and Shephard's output distance function is identical to Farrell's output-oriented efficiency measure.)

The line AA' is the isocost line with a slope equal to the (negative) ratio of the input prices measured for the input represented on the vertical axis and the input represented on the horizontal axis. The isocost line is tangent to the point Q' on the unit isoquant. Measures based on costs can then be established. Technical efficiency E_1 can be measured by the ratio of costs at the frontier point Q and the observed costs, i.e. OQ/OP again, keeping the observed output level and input prices. Farrell introduced a new measure of *Overall Efficiency OE* measured as OR/OP, and the measure *Allocative Efficiency AE* or *Price Efficiency* by OR/OQ. The latter measure shows the relative cost reduction of moving from a frontier point to a point where input costs for the given output is minimised. As we can easily see, the overall efficiency measure decomposes into the product of the technical efficiency and the allocative efficiency; $OE = E_1 \times AE$.

The three cost concepts we need to link the measures to a general situation with multiple outputs and inputs and a transformation relation for the production relationships are

$$
\begin{aligned}
C_j^0 &= \sum_{i=1}^{m} q_i x_{ij}^0 \\
C_j^Q &= \sum_{i=1}^{m} q_i \theta_j x_{ij}^0 \qquad (4) \\
C_j^* &= \min \sum_{i=1}^{m} q_i x_{ij} \text{ subject to } F(Y_j, X_j) = 0, Y_j \geq Y_j^0
\end{aligned}
$$

Here the first cost concept is observed input outlays with q_i as price of input i (common to all units for simplicity) and x_{ij}^0 unit j's observed quantity of input i, C_j^Q is the input costs at the frontier point Q adjusting the inputs with the efficiency score θ_j, and C_j^* is the minimized costs at point Q' given that output is greater or equal to the observed output. The technical input-oriented efficiency can be written $E_1 = C_j^Q / C_j^0$, allocative efficiency is $AE = C_j^* / C_j^Q$ and overall efficiency is $OE = C_j^* / C_j^0$. We can easily see the same multiplicative decomposition of overall efficiency as above.

Notice that it is the analyst that introduces economic behaviour of cost minimisation, not necessarily any of the units under study. The optimised cost C_j^* is thus only a normative benchmark, and does not necessarily reflect behaviour of any observed unit. This point may easily be misunderstood in the literature: it is the analyst that introduces an artificial unit that obeys the frontier technology and operates without any technical inefficiency. Point Q' in Fig.1 serves as a hypothetical benchmark.

Farrell (1957, pp. 260-261) puts forward arguments for weaknesses with allocative or price efficiency and recommends focusing on technical efficiency. His concerns are that price efficiency is sensitive to introduction of new observations through the impact on curvature of isoquants or errors in measurement of prices, and that the current choice of input proportions may be based on past or expected future prices and not on current prices, and will therefore only provide a good measure in a completely static situation.

Estimation methods

Knowledge about a frontier technology can be obtained in two ways; based on engineering information about the production activity in question, or based on observed best practice. The latter was recommended by Farrell (1957) and is usually followed in empirical applications.

Farrell (1957) proposed a method for estimating a best-practice frontier by enveloping the data by a non-parametric piecewise linear function, imposing convex negatively-sloped isoquants and constant returns to scale (CRS). He comments that this way of estimating a production function may not be the best if estimating a frontier is all that is required, but that "it was chosen simply as providing the best measure of technical efficiency" (Farrell, 1957, p. 262).[4] His estimation method in his original Diagram 2 is illustrated in Figure 2, where input coefficients are measured along the axis as in Diagram 1. The CRS frontier production function is estimated by piecewise linear segments spanned by the best-practice observations. In order to keep the end parts of the unit isoquant parallel with the axis and have the type of convex isoquants he wanted, he entered artificial units of value (∞, 0) (a large number suffices in actual applications) for input coefficients in principle for each axis. (We will return to this practice in Section 5.) Farrell (1957) applied his approach in the case of a single output only. His method of estimating the efficiency scores was based on solving a system of linear

[4] He then goes on to suggest statistical approaches to estimating parametric frontier functions that were followed up in the economics literature during the first three decades after his seminal publication (Førsund and Sarafoglou 2002; 2005).

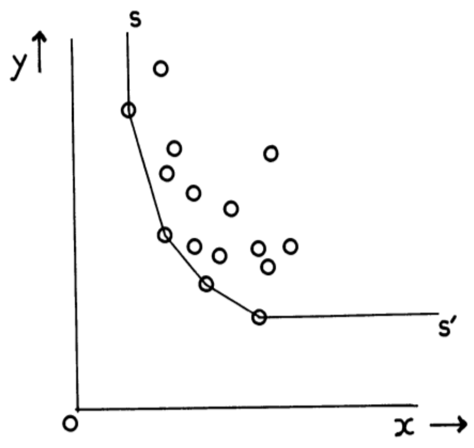

DIAGRAM 2.

Figure 2. The original Farrell (1957) illustration of estimation procedure

equations, stating that observations must lie above to the right in Fig. 2 or on the frontier segments; no points can be located closer to the origin. As mentioned in Section 1, in the discussion of Farrell's 1957 paper at the meeting of the Royal Statistical Association in 1956, Hoffman made the crucial intervention that a newly developed technique, linear programming, could be applied. In Farrell and Fieldhouse (1962) linear programming was applied for the first time to the efficiency problem, however, still restricted to constant returns to scale and a single output. Farrell and Fieldhouse (1962) suggested generalisations both to variable returns to scale and to multiple outputs, but were not completely successful in doing this (Førsund et al. 2009). A group of agricultural economists at Berkeley formalised more successfully the Farrell and Fieldhouse approach and extended the linear programming to multiple outputs (Boles 1967; 1971) (for more references to works by the Berkeley group, see Førsund and Sarafoglou 2002; 2005).

The ratio form of the efficiency measure

Charnes et al. (1978) start out declaring that they want to relate their ideas about efficiency measurement to development in economics by making "reference to production functions and related concepts…" (p. 430). However, they also want to relate their ideas to engineering, and this is actually the starting point for their efficiency definition; it is (allegedly) based on how efficiency is defined within that discipline. The key quote from the engineering literature is: "efficiency is the ratio of the actual amount of heat liberated in a given device to the maximum amount that could be liberated by the fuel" (p. 430). (The reference for the quote is to Encyclopaedia Americana.) This quote leads to the introduction of the ratio approach of maximising a ratio of weighted outputs on weighted inputs subject to this ratio being less than or equal to one for all units, where the weights are the endogenous variables to be determined. This ratio definition is only valid for constant returns to scale. The definition of efficiency for a unit j_0 (using the original symbols in Charnes et al., 1978) is:

$$
\begin{aligned}
&\text{Max}\, h_{j_0} = \frac{\sum_{r=1}^{s} u_{rj_0} y_{rj_0}}{\sum_{i=1}^{m} v_{ij_0} x_{ij_0}} \\
&\text{subject to} \\
&\frac{\sum_{r=1}^{s} u_{rj_0} y_{rj}}{\sum_{i=1}^{m} v_{ij_0} x_{ij}} \le 1 \quad j = 1 \ldots, j_0, \ldots n \\
&u_{rj_0}, v_{ij_0} \ge 0 \ \forall r, i
\end{aligned}
\qquad (5)
$$

In (5) h_{j0} is the efficiency measure, y and x are the output and input vectors, respectively, with s outputs and m inputs, number of units are n, and u_{rj0} , v_{ij0} are the weights associated with outputs and inputs, respectively.[5]

The first ratio expression in (5) defining efficiency in Charnes et al. (1978) is what is termed productivity in economics. The construction of productivity indices constitutes a well-known aggregation problem in economics. A productivity index is closely related to an efficiency index. If a productivity index for a unit is compared to the productivity index of the most productive unit by forming a ratio, then this ratio is an efficiency index using the most productive unit as a benchmark (see Section 4). This is just what the constraints on the productivity index imply in the Charnes et al. (1978) definition.

Conclusions

Farrell (1957) had a production function in mind when defining technical efficiency. A pertinent observation is that the Farrell (1957) definition of efficiency does not depend on the

[5] Later in the paper, in order to avoid the weights turning out zero, the sum of slacks is added in the objective function of multiplier weighted by the inverse of a non-Archimedean infinitely large number M. In Charnes et al. (1979) it is stated that on dual form this addition implies that $\varepsilon = 1/M$ is used as the lower limit for the weights in (5). However, since this number can be arbitrarily close to zero, for practical purposes this may leave economic rates at zero or infinity and the construct is hardly in use anymore in applied studies. There are also more formal reasons for not using the non-Archimedean (Podinovski, 2004b).

method by which the frontier and efficiency scores are estimated, unlike starting from the ratio definition that is usually estimated using linear programming (although formulated as a fractional programming problem). It is also worthwhile to note that the ratio definition as a term is awkward when specifying variable returns to scale, and the name ratio approach is actually not used in Banker et al. (1984), p. 1085, when the ratio problem is derived from the dual to the "envelopment problem", i.e., the optimisation problem leading simultaneously to the estimate of a frontier production function and efficiency scores in the output – input space.

As the definition stands this approach seems to be quite different from the seminal approach of Farrell (1957). We do not see the need for a specific orientation (input or output as in (2) and (3)), and we do not see the proportional scaling that is central to the Farrell definitions. But Charnes et al. (1978), as promised, manage to relate the analysis to economics by turning the fractional program of the ratio definition (5) into an equivalent linear program[6], and furthermore, to show that the generalised optimisation problem of Farrell and Fieldhouse (1962) stated in output and input variables, is in fact the dual to the transformed ratio problem. The estimates of efficiency scores of the different approaches are therefore identical for unique solutions due to the duality property of a linear programming problem. It may then be said that is does not matter which definition that is adopted.

3. The interpretation of shadow prices

Introduction

As mentioned in the Introduction, the term Data Development Analysis (DEA) was introduced in Charnes et al. (1978) and now stands for the model used to estimate efficiency measures like (2) and (3). Due to the standard assumptions imposed on the transformation function $F(X,Y)$ in the second expression in (1) the optimisation problem will be a linear one and we have a primal problem and a dual problem. It is now most common to call the formulation implying enveloping the data like the illustration in Fig. 2 the primal problem. The dual problem follows from the theory of linear programming. The interpretation of the

[6] It may be noted that the latter fractional programming problem, as far as we know, is not in use for practical computations. Since the problem has the Farrell problem as its dual, one may say that starting with the fractional problem (5) is done just to state a definition of efficiency, but as actual computation is concerned it is a detour.

dual variables (also termed weights, multipliers or shadow prices, the last expression being common in Economics) is of key importance for understanding the DEA efficiency model. Unfortunately such an understanding is usually not explicitly found in papers on DEA.

The DEA non-parametric efficiency model

The DEA non-parametric efficiency model has the following production possibility set in the case of variable returns to scale (VRS) (Banker et al., 1984):

$$T = \left\{ (Y, X) : Y \leq \sum_{j=1}^{n} Y_j \lambda_j, X \geq \sum_{j=1}^{n} X_j \lambda_j, \sum_{j=1}^{n} \lambda_j = 1, \lambda_j \geq 0 \right\} \quad (6)$$

For this equation we assume that the axioms of convexity and free disposability for the production possibility set presented in Banker et al. (1984) hold. The scalar variables λ_j are called "intensity weights" in the DEA literature. All values are constrained to be non-negative, and at least one output, one input and one intensity weight have to be strictly positive. Dropping the condition that the intensity weights sum to 1 constant returns to scale is imposed (an assumption of *T* being a cone is introduced).

The set *T* is a polyhedral set with the points $(\sum_{j=1}^{n} Y_j \lambda_j, \sum_{j=1}^{n} X_j \lambda_j)$ representing all possible combinations within the set based on n units of observation. Vertex points will be the efficient points spanning the border of the production possibility set. The weights λ_j have to be estimated in a way consistent with the axiom of the set being the closest convex set enveloping all data points. When we have the optimal solution for these weights we also have an estimate of the production possibility set and the border of the set; commonly termed the frontier production function.

When using LP for both estimating the frontier and the efficiency measures, as done for the first time in Farrell and Fieldhouse (1962), and generalised to multiple outputs and made accessible to the research community in Charnes et al. (1978), then we have the fundamental relationship between a primal solution and a dual solution of an optimal solution. In a technical sense one may then say that whether efficiency is defined using the primal or dual does not matter. However, one should not forget that the basic definition of an efficiency measure in economics is based on the frontier production function concept and formulated in input – output space. It is therefore natural, at least for economists, to view the problem called the envelopment problem in operations research for the primal model and the problem formulated in a shadow price space for the dual problem (the multiplier problem in OR literature).

The standard primal problem (the envelopment problem) in contemporary DEA literature using the Banker et al. (1984) model based on the set T in (6) to estimate Farrell technical efficiency scores for a unit j_0 in the case of variable returns to scale and input orientation of the efficiency measure is[7]:

$$\begin{aligned}
&\text{Min } \theta_{j_0} \\
&\text{subject to} \\
&\sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_{j_0} x_{ij_0}, i = 1,...,m \\
&\sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{rj_0}, r = 1,...,s \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j \ge 0 \text{ , } \theta_{j_0} \text{ sign free}
\end{aligned} \tag{7}$$

We here see the Farrell proportional (or radial) scaling of inputs as in (2) in order for an inefficient point to be projected to the frontier.

The dual problem to problem (7) is

$$\begin{aligned}
&\text{Max } (\sum_{r=1}^{s} u_{rj_0} y_{rj_0} - u_{j_0}) \\
&\text{subject to} \\
&\sum_{i=1}^{m} v_{ij_0} x_{ij_0} = 1 \\
&\sum_{r=1}^{s} u_{rj_0} y_{rj} - \sum_{i=1}^{m} v_{ij_0} x_{ij} - u_{j_0} \le 0, j = 1...j_0...n \\
&v_{ij_0}, u_{rj_0} \ge 0, u_{j_0} \text{ sign free}
\end{aligned} \tag{8}$$

The variables $v_{ij_0}, u_{rj_0}, u_{j_0}$ are the shadow prices on the constraints in (7), which are the input constraints, the output constraints and the convexity constraint, respectively. In DEA it is

[7] Notice that we do not use an infinitesimal small constant ε (a non-Archimedean quantity) explicitly in the DEA models, since we suppose that each model is solved in two stages in order to separate efficient and weakly efficient units (Cooper et al. 2006).

more common to call these variables for multipliers or weights. We then have the fundamental duality result for a unique optimal solution: $\theta_{j_0} = \sum_r u_{rj_0} y_{rj_0} - u_{j_0}$. In addition to the weighted sum of outputs, expressed in dimensionless efficiency measure units, there is the shadow price u_{jo} on the convexity constraint in the case of variable returns to scale.

The interpretation of dual variables will be based on formulating the Lagrangian function for the primal problem. Setting up the Lagrangian for the constrained optimisation problem (7) (as a maximisation problem for convenience) for unit j_0 we have

$$
\begin{aligned}
L &= -\theta_{j_0} \\
&-\sum_{r=1}^{s} u_{rj_0}\left(y_{rj_0} - \sum_{j=1}^{n} \lambda_j y_{rj}\right) \\
&-\sum_{i=1}^{m} v_{ij_0}\left(\sum_{j=1}^{n} \lambda_j x_{ij} - \theta_{j_0} x_{ij_0}\right) \\
&-u_{j_0}\left(\sum_{j=1}^{n} \lambda_j - 1\right)
\end{aligned}
\qquad (9)
$$

The necessary first-order conditions for a solution to problem (9) are:

$$
\begin{aligned}
&\frac{\partial L}{\partial \lambda_j} = \sum_{r=1}^{s} u_{rj_0} y_{rj} - \sum_{i=1}^{m} v_{ij_0} x_{ij} - u_{j_0} \le 0 \ (= 0 \text{ for } \lambda_j > 0),\ j = 1,\ldots, j_0, \ldots, n \\
&\frac{\partial L}{\partial \theta_{j_0}} = -1 + \sum_{i=1}^{m} v_{ij_0} x_{ij_0} = 0 \\
&u_{rj_0} \ge 0 \ (= 0 \text{ for } y_{rj_0} < \sum_{j=1}^{n} \lambda_j y_{rj}), r = 1,\ldots, s \\
&v_{ij_0} \ge 0 \ (= 0 \text{ for } \sum_{j=1}^{n} \lambda_j x_{ij} < \theta_{j_0} x_{ij_0}), i = 1,\ldots, m
\end{aligned}
\qquad (10)
$$

From the first condition we have that the intensity weight λ_{j_0} will be zero for unit j_0 if this unit is inefficient ($\theta_{j_0} < 1$); using the duality result of equality of the two objective functions in (7) and (8) we have $\sum_r u_{rj_0} y_{rj_0} - u_{j_0} < \sum_i v_{ij_0} x_{ij_0} = 1$, i.e. the first condition in (10) holds with strict inequality. The second condition will hold with equality since the efficiency score is unrestricted. We have that a non-positive value of the efficiency score is not admissible under the assumption of at least one output and one input being strictly positive, and at least one intensity weight must be positive. Furthermore, the efficiency score cannot exceed one in the optimal solution; inputs must be scaled down for inefficient units and remain the same for efficient units due to the nature of the minimisation problem. From the two last complementary slackness conditions we have that the shadow prices become zero for variables where we have non-negative slacks, i.e. the units in question are inefficient in these dimensions.

If we have a unique solution to problem (7) then the shadow prices of the output and input constraints can be interpreted by applying the Envelope Theorem for an inefficient unit. However, we know that there may typically be multiple solutions, especially for shadow prices. We will therefore assume that for inefficient units with the projection point to the frontier being in the relative interior of a face we have unique solutions for the endogenous variables. Thus, considering unit j_0, assuming we have an optimal solution to problem (7), we get:

$$\begin{aligned} &\frac{\partial(-\theta_{j_0})}{\partial x_{ij_0}} = \frac{\partial L}{\partial x_{ij_0}} = v_{ij_0}(\theta_{j_0} - \lambda_{j_0}) = v_{ij_0}\theta_{j_0} \Rightarrow \frac{\partial \theta_{j_0}}{\partial(\theta_{j_0} x_{ij_0})} = -v_{ij_0}, i = 1,\dots,m \\ &\frac{\partial(-\theta_{j_0})}{\partial y_{rj_0}} = \frac{\partial L}{\partial y_{rj_0}} = -u_{rj_0}(1 - \lambda_{j_0}) = -u_{rj_0} \Rightarrow \frac{\partial \theta_{j_0}}{\partial y_{rj_0}} = u_{rj_0}, r = 1,\dots,s \end{aligned} \tag{11}$$

Concerning a change in an exogenous variable we have in general that the shadow price on the constraint in question for an inefficient unit measures the impact on the objective function, the efficiency score, of a marginal increase in the variable in question.

Looking at a change in input i for unit j_0 we have that $\lambda_{j_0} = 0$ for unit j_0 being inefficient from the first necessary condition in (10). In the first (direct) interpretation of (11) the shadow price on the constraint is weighted with the efficiency score. But because this latter is a constant, we can evaluate the impact of a change in the input constraint by evaluating the change at the input value $(\theta_{j_0} x_{ij_0})$ that is on the frontier. The unit of measurement for the shadow price weighted with the efficiency score is efficiency score units per measurement unit of the input variable in question. The efficiency score is reduced when an input of an inefficient unit increases. Notice that the unit of measurement of the dual variable v_{ij_0} itself is the inverse of the input unit of measurement.

For an inefficient unit the shadow price on the output constraint is directly interpreted as the increase in the efficiency score of a marginal increase in the output variable in question evaluated at a frontier point. Remember that by definition of the input-oriented efficiency measure the output component of unit j_0 is on the frontier. The unit of measurement for the shadow price is efficiency score units (dimensionless) per measurement unit of the output variable in question.

In the DEA literature the product of a shadow price and an input, $v_{ij_0}x_{ij_0}$ is called a virtual input, and similarly for the product of a shadow price and an output $u_{rj_0}y_{rj_0}$ is called virtual output. As seen from the objective function of the dual problem (8) a virtual output is just expressing the contribution to the efficiency score at the optimal solution of the variable in question. However, as seen from the first constraint in (8) a virtual input is giving its contribution to the share 1 and is dimensionless (remember the measurement unit of the dual variable stated above). To get the contribution to the efficiency score we have to multiply the dual variable with the efficiency score as seen from the first expression in (11).

The interpretation of the shadow prices in terms of standard production function concepts can straightforwardly be made utilising the dual problem (8). The second constraint will hold with equality in an optimal solution for unit j_0 and is the equation of the hyperplane of the corresponding face (called a facet if the face is of full dimension $m + s - 1$). Assuming that unit j is an efficient unit we have by differentiation of the first equation below:

$$
\begin{aligned}
&\sum_{r=1}^{s} u_{rj_0} y_{rj} - \sum_{i=1}^{m} v_{ij_0} x_{ij} - u_{j_0} = 0 \\
&u_{rj_0} dy_{rj} - v_{ij_0} dx_{ij} = 0 \Rightarrow \frac{dy_{rj}}{dx_{ij}} = \frac{v_{ij_0}}{u_{rj_0}}
\end{aligned}
\qquad (12)
$$

This is the economic concept of the marginal productivity of input i in terms of the output of type r evaluated at a point on the face where the projection of the inefficient unit j_0 is located. Using also the types r' and i' of outputs and inputs, respectively, we develop in the same way the following expressions:

$$
\begin{aligned}
&u_{rj_0} dy_{rj} + u_{r'j_0} dy_{r'j} = 0 \Rightarrow \frac{dy_{rj}}{dy_{r'j}} = -\frac{u_{r'j_0}}{u_{rj_0}} \\
&-v_{ij_0} dx_{ij} - v_{i'j_0} dx_{i'j} = 0 \Rightarrow \frac{dx_{ij}}{dx_{i'j}} = -\frac{v_{i'j_0}}{v_{ij_0}}
\end{aligned}
\qquad (13)
$$

The first ratio expression is the marginal rate of transformation between output r and r' and the second ratio expression is the marginal rate of substitution between input i and i'. The faceted form of the frontier production function implies that these three fundamental economic concepts are constant on a face and varies from face to face.

We know that vertex points are extreme-efficient units (efficient units with zero slacks), so the solution for shadow prices for these units will necessarily not be unique because these units belong to more than one face on the surface of the frontier production function. Because the constraint qualification is not satisfied for the vertex units we cannot use the Envelope Theorem for investigating impacts of change in data for such units.

In the DEA literature shadow prices in the meaning of the dual variables v and u have sometimes been used synonymous to market prices. But we must be careful here. Assuming that units located at the frontier follow economic strategies of either minimising costs or maximising profit, we have that ratios of market prices correspond to the ratios of dual variables in (12) and (13) with input price substituting for a corresponding dual variable of an input constraint and an output price for a corresponding dual variable of an output constraint. But notice that without more information we cannot establish a correspondence between a dual variable and a price. Furthermore, we cannot say much about the economic adaptations of an inefficient unit; giving this unit rational objectives of cost minimisation or profit maximisation some explaining has to be done to measure such a unit as inefficient.

It follows from (12) and (13) that the economic concepts of marginal productivity and rates of transformation and substitution cannot be made when the dual variables are zero for inputs or outputs on the face in question; we get values of zero or (plus/minus) infinity. All the economic concepts have economic meaningful values for a face of full dimension only. Furthermore, a zero dual variable for an output in problem (8) means that this variable does not contribute to the efficiency score, and a zero variable for an input means that this input does not contribute to the summing-up condition to unity (the roles will be reversed for an output-oriented problem).

Calculation of shadow prices

There is an interest in the efficiency literature for calculating shadow prices. There is an extensive review in Zhou et al. (2014). An early paper is Färe et al. (1993), using the relationships in (13) on an estimated parametric distance function. However, there are some pitfalls with the procedure of finding shadow prices. As explained above the shadow prices are evaluated at a frontier point. But as stated in Section 2 there is no assumptions about economic behaviour in DEA. So the projection point of an inefficient unit and the evaluation of shadow prices at this point are done by the analyst, and there is by assumption no observed unit doing optimisation. The units turning out being efficient are mostly located at vertex points, but the shadow prices are not unique here, as mentioned above. Furthermore, if the manager of a unit should have a sudden revelation about the frontier and how to realise a frontier point, the projection point is just one of very many points that the manager wants to move to due to his new insights. Sometimes projection points are used as targets for the units, but this idea suffers from the same weakness that a projection point is just one of many possible points to try to reach. A radial projection point does not necessarily serve an objective function of e.g. maximising profit.

As an example let us assume that we have just one output and two inputs and that point Q' in Fig. 1 corresponds to a corner point in Fig. 2. If we know the price q_2 of one input we can then try to estimate the other price q_1 that is unknown by using the second relation in (13) introducing the condition for cost minimisation: $q_1 = q_2 v_1 / v_2$. However, at a corner point the shadow prices are not known, so the left-hand or right-hand values have to be used. As is evident from Fig. 2, assuming the price q_2 to be known, we get varying estimates for q_1 for different corner points acting as cost-minimising benchmarks for the observations.

Conclusions

The concept of shadow prices is generally used by economists as the name for Lagrangian parameters used for constraints in an optimisation problem when a Lagrangian is employed. Using the DEA model of Banker et al. (1984) the constraints in the LP problem set up to estimate efficiency scores comprise outputs, inputs and securing convexity. It is vital for the understanding of the solution that the interpretations of the shadow prices (also called weights or multipliers) are clear. It is often the case in the DEA literature that these interpretations are not explicitly stated. The objective function is the efficiency score for input orientation (or the inverse of the score for output orientation). The shadow prices in an optimal solution therefore express the change in the objective function, the efficiency score, by a change in the constraint caused by a change in the exogenous output or input observation. Evaluation of the derivatives of the value function (unique optimal solutions inserted in the objective function) the Envelope Theorem yields that the shadow prices express the marginal change in the efficiency score for an inefficient unit, with a frontier projection point in the interior of a face, of marginal changes in outputs and inputs, respectively. This gives a link to economic concepts such as marginal rates of transformation between outputs, marginal rates of substitution between inputs, and marginal productivities. All these concepts are expressed by ratios of the appropriate shadow prices. Using economic conditions for optimality deriving so-called shadow prices is problematic. One problem is that a corner point does not have any unique solution for shadow prices, and another problem is that having to know one input or output price the estimate of other prices will vary with the observation, probably in conflict with the standard assumption of give common prices for all the units.

4. Productivity interpretations of the Farrell efficiency measures

Introduction

The ratio formulation of the efficiency measure in (5) due to Charnes et al. (1978) leads to the measure being interpreted as a productivity measure with the multipliers or shadow prices as weights. However, this may be an awkward interpretation and may have been the motivation behind the interest of imposing restrictions on the weights to get a "sounder" or more acceptable expression for productivity. But we will show that Farrell measures of productivity like the measures in (2) and (3) can be given a straightforward productivity interpretation without involving the shadow prices.

The Farrell suite of efficiency measures

Farrell (1957) defined two technical measures of efficiency, the input-oriented measure based on scaling inputs of inefficient units with a common scalar, projecting the point radially to the frontier keeping observed output constant, and the output-oriented measure scaling outputs of inefficient units with a common scalar, projecting the point radially to the frontier keeping observed inputs constant. The measures were defined for a frontier function exhibiting constant returns to scale. However, he also discussed variable returns to scale and studied this further in Farrell and Fieldhouse (1962) without explicitly introducing measures reflecting scale properties. This was done in Førsund and Hjalmarsson (1974) and (1979), developing a family of five efficiency measures. This was illustrated using a smooth variable returns to scale frontier production function exhibiting an *S*-shaped graph as typical for neoclassical production functions obeying the *Regular Ultra Passum Law* of Frisch (1965)[8]. This may be the reason for this family of efficiency measures being rather unknown in the DEA literature. However, the efficiency measures are valid for any type of frontier function as long as a basic requirement of the variation of the elasticity of scale is fulfilled. It is in particular valid for the generic DEA model exhibiting variable returns to scale (VRS) in Banker et al. (1984) that will be used in this paper.

The family of Farrell efficiency measures is illustrated in Figure 3 (Førsund and Hjalmarsson, 1979) in the case of the frontier within a non-parametric framework being a piecewise linear

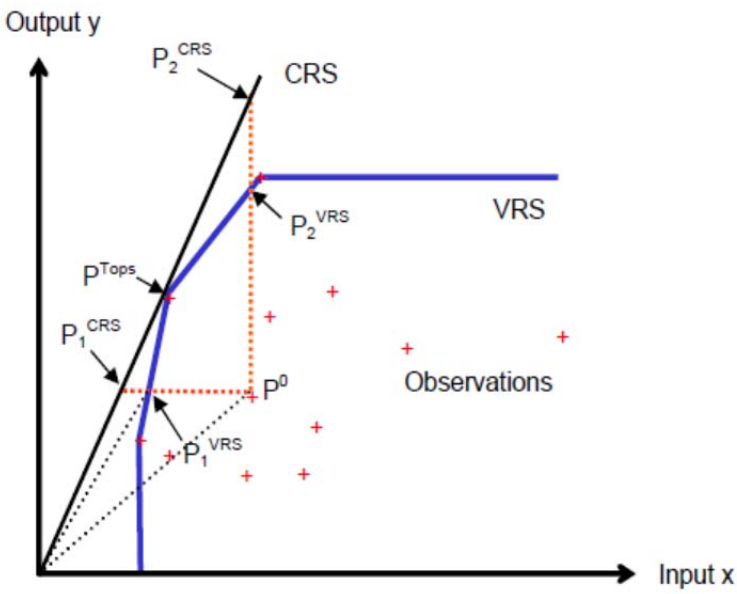

Figure 3. The Farrell efficiency measures applied to a piecewise linear frontier

function. The point of departure is the observation $P^0 = (y^0, x^0)$ that is inefficient with respect to the VRS frontier. The reference point on the frontier for the input-oriented measure E_1 with respect to the VRS frontier is $P_1^{VRS} = (y^0, x_1^{VRS})$, and the reference point on the frontier for the output-oriented measure E_2 with respect to the VRS frontier is $P_2^{VRS} = (y_2^{VRS}, x^0)$. A second

[8] The Regular Ultra Passum Law requires that the scale elasticity decreases monotonically from values greater than one, through the value one to lower values when moving along a rising curve in the input space.

envelopment is indicated by the ray from the origin being tangent to the point P^{Tops}. (We will return to the interpretation of this point below.) This frontier exhibits constant returns to scale (CRS). The reference points on the frontier are $P_1^{CRS} = (y^0, x_1^{CRS})$ and $P_2^{CRS} = (y_2^{CRS}, x^0)$. The dotted factor ray from the origin to the observation gives the productivity of the observation, and the dotted factor ray from the origin to a reference point on the VRS frontier gives the productivity of this reference point. As is easily seen from Fig. 3 the productivity at the CRS envelopment is the maximal productivity obtained on the VRS frontier. Comparing the observation with the reference point $P^{Tops} = (y^T, x^T)$ therefore gives the relative productivity of an observation to the maximal productivity on the VRS frontier. Continuing Farrell's numbering of measures a measure E_3 is introduced covering this measurement and is therefore termed the measure of *technical productivity*.[9] The two remaining efficiency measures E_4 and E_5 introduced in Førsund and Hjalmarsson (1979) are the scale efficiency measures[10] comparing the productivity of the reference points P_1^{VRS} and P_2^{VRS} respectively with the point P^{Tops} of maximal productivity on the frontier.

All Farrell measures of efficiency can be given an interpretation of relative productivity; the productivity of the observation relative to specific points on the VRS frontier. Before showing the relative productivity interpretation in the case of a single output and a single input in a general setting, let us state the definitions of the Farrell input-and output-oriented technical efficiency measures, starting with the definition of the production possibility set (1a) in Section 2. By assumption let the set T exhibit variable returns to scale (VRS) of its frontier. The input-and output-oriented efficiency measures can be defined as

$$\begin{aligned} E_1(y,x) &= Min_\mu \{\mu : (\mu x, y) \in T\} \\ E_2(y,x) &= Min_\lambda \{\lambda : (x, y/\lambda) \in T\} \end{aligned} \tag{14}$$

The relative productivity interpretation can be shown in the following way using Fig. 3, starting with the input-oriented efficiency measure:

$$\frac{y^0/x^0}{y^0/x_1^{VRS}} = \frac{y^0/x^0}{y^0/E_1x^0} = E_1 \tag{15}$$

The same productivity interpretation holds for the output-oriented efficiency measure:

$$\frac{y^0/x^0}{y_2^{VRS}/x^0} = \frac{y^0/x^0}{(y^0/E_2)/x^0} = E_2 \tag{16}$$

[9] In Førsund and Hjalmarsson (1979), introducing this measure, it was called the gross scale efficiency.
[10] In Førsund and Hjalmarsson (1979) these measures were called measures of pure scale efficiency.

In the input-oriented case we adjust the observed input quantity so that the projection of the observation is on the frontier, and in the output-oriented case we adjust the observed output, using the symbols for adjusted input and output introduced above.

For the three remaining measures we will make a crucial use of the CRS envelopment in order to calculate the measures. The notation E_1^{CRS} and E_2^{CRS} making explicit reference to the CRS envelopment as the benchmark frontier together with $\mathrm{P}^{\mathrm{Tops}} = (y^{\mathrm{T}}, x^{\mathrm{T}})$ will be used. The measure of technical productivity is

$$\begin{aligned} E_3 &= \frac{y^0 / x^0}{y^T / x^T} = \frac{y^0 / x^0}{y^0 / E_1^{CRS} x^0} = E_1^{CRS} \\ E_3 &= \frac{y^0 / x^0}{y^T / x^T} = \frac{y^0 / x^0}{(y^0 / E_2^{CRS}) / x^0} = E_2^{CRS} \Rightarrow E_3 = E_1^{CRS} = E_2^{CRS} \end{aligned} \tag{17}$$

The first expression in each of the two lines of the equations is the definition of the measure of technical productivity using the productivity at the point $\mathrm{P}^{\mathrm{Tops}}$ as a reference. The second expressions, input-orientation or output-orientation, respectively, show the most convenient way of calculating the productivity measure. Using the CRS envelopment the maximal productivity for the VRS technology is the same along the entire ray from the origin going through the point $\mathrm{P}^{\mathrm{Tops}}$. The productivity measure E_3 is equal to both the input-oriented measure and the output-oriented measure using the CRS envelopment as the frontier. It is easy to see geometrically that in the case of using the CRS envelopment the two efficiency measures must be identical, as pointed out by Farrell (1957).

Measures for scale efficiency are also defined using a relative productivity comparison. The input-oriented scale efficiency E_4 (keeping output fixed) and the output-oriented scale efficiency E_5 (keeping input fixed) are:

$$\begin{aligned} E_4 &= \frac{y^0 / x_1^{VRS}}{y^T / x^T} = \frac{y^0 / E_1 x^0}{y^0 / E_1^{CRS} x^0} = \frac{E_1^{CRS}}{E_1} = \frac{E_3}{E_1} \\ E_5 &= \frac{y_1^{VRS} / x^0}{y^T / x^T} = \frac{(y^0 / E_2) / x^0}{y^0 / E_2^{CRS} x^0} = \frac{E_2^{CRS}}{E_2} = \frac{E_3}{E_2} \end{aligned} \tag{18}$$

The relative productivity comparison for input-oriented scale efficiency in Fig. 1 is between the observed output on the efficiency-corrected input on the VRS frontier and the maximal productivity at the $\mathrm{P}^{\mathrm{Tops}}$ - point = $(y^{\mathrm{T}}, x^{\mathrm{T}})$. For output-oriented scale efficiency we have an analogous construction. The calculations of the scale efficiency measures can either be based on the ratios between the efficiency scores for input-oriented efficiency relative to the VRS frontier and the CRS envelopment or expressed as deflating the technical productivity measure with the relevant efficiency measures relative to the VRS frontier.

The concepts of elasticity of scale and technically optimal scale

Before generalising the relative productivity interpretation to multiple outputs and inputs we need to introduce the concept of elasticity of scale. The definition of scale elasticity for a frontier production function is the same whether it is of the neoclassical differential type $F(y,x)=0$ or if the production possibility set has a faceted envelopment border like in the DEA case. We are looking at the maximal proportional expansion β of outputs for a given proportional expansion α of inputs, i.e. looking at $F(\beta y,\alpha x)=0$. The scale elasticity is defined as the derivative of the output expansion factor w.r.t. the input expansion factor on the average value of the ratio of the output factor on the input factor[11]:

$$\varepsilon(x,y)=\frac{\partial\beta(x,y,\alpha)}{\partial\alpha}\frac{\alpha}{\beta}=\frac{\partial\beta(\alpha,x,y)}{\partial\alpha}\Big|_{\alpha=\beta=1} \qquad (19)$$

The scale elasticity is evaluated without loss of generality for $\alpha=\beta=1$. In the DEA case with non-differentiable points (vertex points and points on edges) the expression above is substituted with the right-hand derivative and the left-hand derivative, respectively, at such points (Krivonozhko et al 2004; Førsund et al 2007; Podinovski et al 2009; Podinovski and Førsund 2010).

Returns to scale is defined by the value of the scale elasticity; increasing returns to scale is defined as $\varepsilon > 1$, constant returns to scale as $\varepsilon = 1$ and decreasing returns to scale as $\varepsilon < 1$.

For a production function with variable returns to scale there is a connection between the input- and output- oriented measures via the scale elasticity. Following Førsund and Hjalmarsson (1979) in the case of a frontier function for a single output and multiple inputs we have

$$E_2=E_1^{\bar{\varepsilon}}\Rightarrow E_1 \overset{>}{\underset{<}{=}} E_2 \text{ for } \bar{\varepsilon} \overset{>}{\underset{<}{=}} 1, \qquad (20)$$

[11] See Hanoch (1970); Panzar and Willig (1977); Starrett (1977).

where the variable $\bar{\varepsilon}$ is the average elasticity of scale along the frontier function from the evaluation point for the input-saving measure to the output-increasing measure. In Førsund (1996) this result was generalised for multiple outputs and inputs in the case of a differentiable transformation relation $F(y,x)=0$ as the frontier function, using the *Beam variation equations* of Frisch (1965). This result holds for points of evaluation being projection points in the relative interior of faces. The path between the points will be continuous although not differentiable at vertex point or points located at edges.

We must distinguish between scale elasticity and scale efficiency (Førsund 1996). Formalising the illustration in Fig. 3 the reference for the latter is the concept of *technically optimal scale* of a frontier function (Frisch 1965). The set of points $TOPS^T$ having maximal productivities for the border of the set T in (1a) with the frontier exhibiting VRS can be defined as (Førsund and Hjalmarsson 2004)

$$TOPS^T = \{(x,y) \mid \varepsilon(x,y)=1, (x,y)\in T\} \tag{21}$$

It must be assumed that such points exist and that for outward movements in the input space the scale elasticity cannot reach the value of 1 more than once for a smooth neoclassical frontier. However, it can in the DEA case be equal to 1 for points on the same face. The point (y^{T}, x^{T}) used above is now replaced by vectors y^{T} and x^{T} belonging to the set $TOPS^T$. From production theory we know that in general a point having maximal productivity must have a scale elasticity of 1. In a long-run competitive equilibrium efficient production units will realise the technically optimal scale with the scale elasticity of 1 implying zero profit.

The productivity interpretation of the efficiency measures in the general case

The interpretation of the five Farrell measures as measures of relative productivity can straightforwardly be generalised to multiple outputs and inputs. Introducing general aggregation functions $g_y(y_1, y_2,\ldots,y_M)$ and $g_x(x_1, x_2,\ldots, x_N)$ for outputs and inputs, respectively, increasing in the arguments and being homogeneous of degree 1 in outputs and inputs, respectively (y and x are now interpreted as vectors and y_1, x_1 etc. as elements of the respective vectors), we have, starting with the definition of relative productivity in the input-oriented case for an observation (y^0, x^0):

$$\frac{g_y(y^0)/g_x(x^0)}{g_y(y_1^{VRS})/g_x(x_1^{VRS})} = \frac{g_y(y^0)/g_x(x^0)}{g_y(y^0)/g_x(E_1x^0)} = \frac{g_y(y^0)/g_x(x^0)}{g_y(y^0)/E_1 g_x(x^0)} = E_1 \tag{22}$$

In the first expression relative productivity is defined in the input-oriented case using the observed vectors y^0, x^0 and the vectors y_1^{VRS}, x_1^{VRS} for the projection onto the VRS frontier at analogous to the point P_1^{VRS} in Fig. 3 in the two-dimensional case. In the second expression the vectors for y_1^{VRS} and x_1^{VRS} are inserted, keeping the observed output levels y^0 and contracting the observed input vector using the input-oriented efficiency E_1 to project the inputs x^0 to the VRS frontier. In the third expression the homogeneity property of the input index function is used.

In the case of output orientation of the efficiency measure E_2 we get in the multiple output – multiple input case following the procedure above:

$$\frac{g_y(y^0)/g_x(x^0)}{g_y(y_2^{VRS})/g_x(x_2^{VRS})}=\frac{g_y(y^0)/g_x(x^0)}{g_y(y^0/E_2)/g_x(x^0)}=\frac{g_y(y^0)/g_x(x^0)}{(g_y(y^0)/E_2)/g_x(x^0)}=E_2 \tag{23}$$

Using the general aggregation functions $g_y(y)$, $g_x(x)$ the measure of technical productivity can be derived using input- or output-orientation:

$$\begin{aligned}
&E_3=\frac{g_y(y^0)/g_x(x^0)}{g_y(y^T)/g_x(x^T)}=\frac{g_y(y^0)/g_x(x^0)}{g_y(y^0)/g_x(E_1^{CRS}x^0)}=\frac{g_y(y^0)/g_x(x^0)}{g_y(y^0)/E_1^{CRS}g_x(x^0)}=E_1^{CRS}=\\
&E_3=\frac{g_y(y^0)/g_x(x^0)}{g_y(y^T)/g_x(x^T)}=\frac{g_y(y^0)/g_x(x^0)}{g_y(y^0/E_2^{CRS})/g_x(x^0)}=\frac{g_y(y^0)/g_x(x^0)}{(g_y(y^0)/E_2^{CRS})/g_x(x^0)}=E_2^{CRS}\Rightarrow \\
&E_3=E_1^{CRS}=E_2^{CRS}
\end{aligned} \tag{24}$$

We obtain the same relationship between the technical productivity measure and the oriented measures with the CRS envelopment as in the simple case illustrated in Fig.3.

The case of multi-output and –input is done in the same way for the scale efficiency measures as for the other measures utilising the homogeneity properties of the aggregation functions:

$$\begin{aligned}
&E_4=\frac{g_y(y^0)/g_x(x_1^{VRS})}{g_y(y^T)/g_x(x^T)}=\frac{g_y(y^0)/g_x(E_1x^0)}{g_y(y^0)/E_1^{CRS}g_x(x^0)}=\frac{g_y(y^0)/E_1g_x(x^0)}{g_y(y^0)/E_1^{CRS}g_x(x^0)}=\\
&\quad\frac{E_1^{CRS}}{E_1}=\frac{E_3}{E_1}\\
&E_5=\frac{g_y(y_2^{VRS})/g_x(x^0)}{g_y(y^T)/g_x(x^T)}=\frac{g_y(y^0/E_2)/g_x(x^0)}{(g_y(y^0)/E_2^{CRS})/g_x(x^0)}=\frac{(g_y(y^0)/E_2)/g_x(x^0)}{(g_y(y^0)/E_2^{CRS})/g_x(x^0)}=\\
&\quad\frac{E_2^{CRS}}{E_2}=\frac{E_3}{E_2}
\end{aligned} \tag{25}$$

Again, we obtain the same relationship between the technical productivity measure and the oriented measures defining scale efficiency as in the simple case illustrated in Fig. 3. The calculations of the scale efficiency measures can either be based on the ratios between the efficiency scores for input-oriented efficiency relative to the VRS frontier and the CRS envelopment or expressed as deflating the technical productivity measure with the relevant efficiency measures relative to the VRS frontier.

Conclusions

Charnes et al (1978) introduced the ratio form of productivity measures for estimating the efficiency scores via estimating the weights in a linear aggregation of outputs and inputs used to measure the productivity of a unit, and then maximising this productivity subject to no productivity ratio using these weights for all units being greater than one (as a normalisation, see Section 2). This ratio measure is said to be inspired by how efficiency is defined in the engineering literature. However, this way of defining efficiency measures is not as satisfactory for economists as the Farrell approach introducing explicitly a frontier production function as a reference for efficiency measure definitions.

The original Farrell measures developed for constant returns to scale (CRS) can be extended to five efficiency measures for a frontier production function exhibiting variable returns to scale (VRS); input- and output technical efficiency, input- and output scale efficiency, and the technical productivity measure. The relationship between the two measures of technical efficiency involves the average scale elasticity value between the two frontier projection points along the frontier surface. The technical productivity measure and the two scale efficiency measures are developed based on the Frisch (1965) concept of technically optimal scale, predating the use of the concept most productive scale size in the DEA literature with almost 20 years.

It does not seem to be recognised in the DEA literature that in the general case of multiple outputs and inputs the Farrell efficiency measures can all be given productivity interpretations in a more satisfactory way than the ratio form of Charnes et al. (1978). Using quite general aggregation functions for outputs and inputs with standard properties, it has been shown that all five Farrell efficiency measures can be given a productivity interpretation employing a proper definition of productivity. Each of the two technical efficiency measures and the technical productivity measure can be interpreted as the ratio of the productivity of an inefficient observation and the productivity of its projection point on the frontier, using the general aggregation equations. Of course, we have not estimated any productivity index as such, this remains unknown, but that was not the motivation of the exercise in the first place.

5. Weight restrictions in DEA

Introduction

Early empirical application of the DEA analysis of Charnes et al. (1978), based on the ratio definition of efficiency found there, revealed some specific features of the solutions that created alarm. The problem were of two types,

i) Too many of the units under study became efficient
ii) Too many weights in the LP solution of the DEA problem became zero, thus allegedly implying that the corresponding outputs or inputs for units may not count in the solution for the efficiency score.

The proposed solution was to restrict the weights in some ways. Most of the papers concerned with weight restrictions start with the Charnes et al. (1978) efficiency definition of weighted outputs over weighted inputs as in (5). This may explain the interest in the weights. It is also common to associate the weights with value judgements, and state that introducing restrictions on these weights is to introduce values. However, the question is if this is warranted.

Forms of weight restrictions

The earliest applications of weight restrictions were done within a CRS model (Thomson et al. (1986), Dyson and Thanassoulis, 1988). The constraints can also be introduced in a VRS model like (8). The most common ways of restrictions have been imposed on the weights in the dual CRS model will be shown below. The types of weight restrictions can be classified in several ways (see the survey in Allen et al., 1997). We will focus on the most common types of restrictions.

i) Absolute restrictions:

$$\delta_i \le v_i \le \tau_i\ ,\ \rho_r \le u_r \le \eta_r\ , i = 1,..,m\ , r = 1,...,s \tag{26}$$

Bounds are put on the range of the weights. The problem is how to determine such bounds. It is also a question of how many bounds to impose. Remember that the weights are specific to a unit, so if all units are to get restrictions we are talking about $2m \times s \times n$ limits to be set; a rather high number for a typical dataset. So a choice of uniform limits is often done.

The use of experts setting limits is mentioned in the literature. The expert with knowledge of the type of production in question cannot be any expert since he is asked to do this in dual space. It seems difficult to find persons working in industry that are able to associate anything with a dual space. In the few papers that have tried to find weight restrictions by working with sector experts it is difficult to see any generalising principles appearing from the often considerable effort spent (Joro and Viitala 2004).

ii) Relative restrictions, i.e., restrictions on marginal productivities, marginal rates of substitution and marginal rates of transformation (the last two are used in Charnes et al., 1990):

$$\begin{aligned} &\frac{v_i}{u_r} \ge \gamma_i\ , i = 1,..,m\ , r = 1,...,s \\ &l_{1i} \le \frac{v_i}{v_1} \le k_{1i}\ ,\ L_{1r} \le \frac{u_r}{u_1} \le K_{1r}\ ,\ i = 2,\ldots,m\ , r = 2,...,s \end{aligned} \tag{27}$$

One may believe that it is easier to get information on economic ratios, and that sector experts will be more familiar with these. Charnes et al. (1990) also suggest the use of prices for setting limits for ratios. However, this is hardly relevant, because this assumes that the hypothetical unit on the frontier, located at a point in the relative interior of a facet, is actually minimising costs or maximising profit, but this cannot apply to inefficient units, and then certainly not to their projections, and efficient units are typically vertex points that are not differentiable, so it is also without good meaning to appeal to economic optimising conditions for such points. The standard assumption in DEA is, after all, that data are not generated based on economic optimisation.

iii) Share of efficiency score (Wong and Beasley, 1990):

$$\phi_r \le \frac{u_r y_{rj}}{\sum_{r=1}^{s} u_r y_{rj}} \le \psi_r \tag{28}$$

As seen above Wong and Beasley (1990) introduced a special version of weight constraints by constraining the share one variable had of the efficiency score to be within bounds. An exercise using the approach is found in Beasley (1995). Although Wong and Beasley are aware of the multipliers being dimensionless, constraining the shares of the efficiency score is called introducing value judgements (p. 831). On the background of Section 3 it is difficult to agree with such a terminology. Again, the technical role of shadow prices in a programming problem is confused with values in an economic sense.

A critical review of seminal contributions

There are several surveys of weight restrictions in the literature (Allen et al., 1997; Pedraja-Chaparro et al., 1997; Thanassoulis et al., 2004; 2008). However, no examination of the soundness of the approach is offered. Therefore a more critical view of the approach is warranted (see Førsund (2013) for an extensive critical review).

The first (published) paper to raise the weight-restriction issue in DEA is Thompson et al. (1986). The efficiency problem set up involves just six units, sites for locating a high-energy physics lab in Texas, and three inputs are specified assuming the same output for all the units. The problem was that running this DEA model five of the six sites turned out as efficient.[12] The "system task force" started with manipulation of the weights in order to "weight the problem's primary dimensions to establish preference for one site versus another (p. 37)." This was done by imposing restrictions in what they called the price-weight space. The concept of assurance regions to characterise lower and upper bounds for the input multipliers was born.

In Charnes et al. (1990) the problem faced was like the problem in Thompson et al. (1986); too many efficient units. Running a standard DEA model with constant returns to scale on data for banks, DEA even recognised a few "notoriously inefficient banks" as efficient. A more objective assessment of managerial performance was desired. The solution was to base the estimate of the efficiency weights on only a few units declared efficient by bank experts. In the example only three banks, recognised as pre-eminently efficient, were chosen to represent the technology. It was stated that these three banks "were sufficient to provide for a reasonable range of flexibility in relative valuations of inputs and outputs" (p. 75). The ratio

[12] It may be argued that the main reason for this result was the lack of degree of freedom; three variables and six observations.

model of Charnes et al. (1978) was restricted in multiplier space by cones formed by the three efficient units. Imposing the restriction on the cones corresponds to transforming the data for the other units to comply with the shape of the frontier production function determined by those units. Data are transformed using the few efficient units to span the production possibility set, such that standard DEA software can be used after the transformation.

However, to discard information from real data sets and basing the estimation of efficiency measures on a very few units chosen by some experts, does not seem to represent a proper scientific approach to the estimation problem at hand. The evaluation will obviously depend on the few selected units (Charnes et al. 1990, p. 81).

In Dyson and Thanassoulis (1988) there is no mentioning of a production function at all, and neither any reference to Farrell (1957). Their concern is a different one from that in Thompson et al. (1986). They are worried about the complete weight flexibility in the ratio model, since “some DMUs [are] being assessed only on a small subset of their inputs and outputs, while their remaining inputs and outputs are all but ignored” (p. 563). Furthermore, they state (p. 564): “Few would argue against reducing weight flexibility in DEA, since doing so would ensure that the subsequent assessment not only cannot effectively ignore any inputs or outputs, but also would assign weights to inputs and outputs more in line with some general view of their perceived importance.” This statement reveals that they want to attach values beyond a production-function frame of reference to output and inputs.

As to the concern of zero weights influencing the efficiency score it should be noted from (8) that for the input-oriented problem the sum of the product of shadow prices and the corresponding inputs is equal to one, so the weights for inputs do not have a direct impact on the efficiency score, but, of course, indirectly through the influence on the solution for the output weights. (In the case of output orientation it is the sum of the product of shadow prices and outputs that are equal to one and the concern should only be about input weights.)

Regarding zero weights it is important to have in mind that the data at hand actually determine the outcome (Olesen and Petersen 1996). Zero weights may appear because data do not contain sufficient information to avoid this given the variables specified as outputs and inputs. Olesen and Petersen (1996) are very clear on the connection between the data and the resulting form of the frontier production function when using the DEA model to estimate it. They analyse consequences of ill-conditioned data sets in detail, and point out the role of facets of full dimension if estimates of rates of transformation and substitution is also wanted, and not only an efficiency score. There cannot be any zero value for multipliers or weights in the relative interior of a fully dimensional efficient facet. The problem is that variation in data may not support a full set of the rates. A data set is called ill-conditioned if a relatively large number of the units are located in areas where a full set of ratios does not exist.

A typical source of confusion can be found in the following statements (Dyson and Thanassoulis 1988, pp. 564-565): "However, it is difficult to decide exactly how weights are to be constrained within a DEA assessment model in the general case, as weights cannot be readily interpreted", and furthermore: "In general, the weights in a DEA model do not have a clear interpretation, which makes constraining them arbitrary." However, as shown in Section 3 the shadow prices on output and input constraints have, indeed, a well-defined mathematical interpretation as the change in the objective function by a marginal change in the constraints. But this has nothing to do with putting values on outputs and inputs as such.

Some problems with imposing weight restrictions

An interesting new insight in Allen et al. (1997) is, however, presented concerning the nature of the Farrell radial technical efficiency measure when weight restrictions are introduced. For input and output variables that have their marginal rates constrained, it is shown that the radial nature of the Farrell measure is lost if the constraints are binding, even if the equality between the ratio definition and the scaling factor still holds. In case of absolute constraints on weights the equality may also be lost.

The possible divergence between the ratio measure of efficiency of Charnes et al. (1978) and the radial scaling factor of Farrell (1957) due to weight restrictions is given a thorough and extensive treatment in Podinovski and Athanassopoulos (1998) and in a series of related follow-up papers (Podinovski 1999; 2001a; 2001b; 2004b). It is rigorously shown that placing absolute weight restrictions in a DEA model equivalent to the model (5) in Charnes et al. (1978) generally does not lead to the correct evaluation of the relative efficiency of the assessed unit.

We will argue that the most defensible approach to weight restriction is that there is additional information about the shape of the production function. As we have seen in Section 3 marginal productivities and rates of transformation and substitution are expressed by ratios of dual variables. But as is evident from Section 3 these properties are face-specific, so to impose general restrictions seems inappropriate. It would, indeed, be a formidable task to get enough information about properties of each face.

In three related contributions (Podinovski 2004a; Podinovski 2005; Podinovski 2007), Podinovski establishes a way of transforming information about trade-offs between outputs or inputs in input-output space and work out the corresponding restrictions on weights in the dual space. Incorporating trade-off information will extend the production possibility set. A main property of the trade-off approach is then that the technological meaning of efficiency in terms of the radial contraction factor, the Farrell technical efficiency measure, is not changed. It is shown that introducing so-called value judgements for introducing weight constraints in the dual space will not lead to the efficiency measure calculated using the ratio definition being equal to the Farrell technical efficiency measure in Section 2.

There are some problems with the trade-off approach, however. It is underlined in Podinovski (2005) that the trade-offs are not the same as marginal rates of transformation and substitution (however, trade-offs may be regarded as bounds on such rates). Furthermore, the trade-offs are assumed to be valid for all observations. To check if this holds for a realistic data set is, indeed, some task. When estimating a frontier function concept the actual technology applying to each observed unit is not investigated, it is the pure data that are used. The problem of getting more information about the frontier function rates remains unresolved.

The shadow prices appearing in linear programming and occurring in the ratio definition of efficiency are not measures of economic values. If an overall efficiency measure is sought, then the values have to be found in another way, and treated as exogenous to the programming problem, just like the original definition of overall efficiency in Farrell (1957), introducing input prices.

Conclusions

A main motivation for introducing constraints on weights is the occurrence of zero values in the solutions for shadow prices on the constraints. However, zero values is a consequence of positive slacks in the input or output constraints that occur quite often. The problem formulation is such that using weights to express a productivity index becomes awkward. It must not be forgotten that the objective of the benchmarking exercise is to estimate efficiency scores, not productivity indices. The nature of the data combined with the specification of input and output variables may result in some of the variables ending up with zero weights. But trying to fix this by introducing constraints of positive shadow prices implies that the imposed values of the constraints will actually appear in the solution, if the constraints are binding. The analyst may then be unduly dictating the solution and influencing the form of the production possibility set without any real empirical understanding of how this set looks like.

6. Cross efficiency

Introduction

The concept of cross-efficiency evaluation was first proposed by Sexton et al. (1986) and followed up later in Doyle and Green (1994), the latter paper having 296 citations in the Web of science per 31.10.2013[13] reflecting a significant number of applications. The key starting point is the restriction in the original Charnes et al. (1978) paper that the productivity of the unit under investigation, with output- and input weights to be determined solving an optimisation problem, and the productivities of all the other units using the same weights for calculating productivities must be equal to or less than 1 (a normalisation) (see (5) in Section 2). The nature of the optimisation problem of maximising the productivity of the unit under investigation has led to the introduction of the term *self-rated efficiencies* (Sexton et al. 1986) (or *self-appraisal* in Doyle and Green 1994) for determining the weights calculating the Farrell (1957) technical efficiency score, using the formulation that the unit can determine the weights that puts it in the best possible light. According to Doyle and Green (1994) obtaining such a 'simple' efficiency score for a unit can be thought of as a process of *self-appraisal*. To use own weights to calculate the productivity of other units was then called *peer appraisal* in Doyle and Green (1994).

Introducing constraints on weights was introduced in Sexton et al. 1986 at the same time as cross-efficiency was introduced.[14] However, Doyle and Green (1994) advocate that cross-

[13] The paper by Sexton et al. (1986) is not listed in the Web of Science.
[14] However, this pioneering introduction of weight restrictions is not recognised in the literature (see the surveys in Allen et al. 1997; Førsund (2013); Pedraja-Chaparro 1997; Thanassoulis et al. 2004; Thanassoulis et al. 2008). But the critique of weight restrictions in Section 5 also applies to Sexton et al. (1986).

efficiency is superior and state that constraints on weights have something of the air of arbitrariness, and furthermore:

> "We suggest that cross-efficiency, with its intuitive interpretation as peer-appraisal, has less of the arbitrariness of additional constraints, and has more of the right connotations of a democratic process, as opposed to authoritarianism (externally imposed weights) or out and out egoism (self-appraisal)" (Doyle and Green, 1994, p. 570).

The fact that the solutions for weights may not be unique (vertex points will always have multiple solutions) led Sexton et al. (1986) to search for *secondary objectives* to deal with this problem (Liang et al., 2008a; Wang and Chin, 2010a). Minimising the other DMUs' cross-efficiencies was called 'aggressive', while trying to make these cross-efficiencies large was called 'benevolent'.

In the cross-efficiency literature one can often find that the problem of determining weights is portrayed as a choice open to the individual unit, or DMU in DEA lingo. Whether this is the reason or not, game theory concepts have been introduced in the sense that the DMUs are engaged in a strategic game with other units when determining weights.

Formal definition of cross-efficiency

The starting point for cross-efficiency models is the Charnes et al (1978) (CCR) ratio model (5) for n units employing m inputs x to produce s outputs y with weights v and u for inputs and outputs, respectively, calculating an efficiency measure h_{j0} for unit j_0 (see Section 1 on Efficiency measure definitions). The left-hand side of the constraint in (5) is defined as cross efficiency scores using the weights of unit j_0 under investigation as weights for a unit j (j =1,..,n). Expanding the unit j_0 under investigation to all units a matrix E of cross efficiencies is formed, for simplicity changing notation to the one used by Doyle and Green using k instead of j_0 for the unit under investigation:

$$E_{kj} = \frac{\sum_{r=1}^{s} u_{rk} y_{rj}}{\sum_{i=1}^{m} v_{ik} x_{ij}} \leq 1,\ k, j = 1,\ldots,n \qquad (29)$$

In (29) k is the row index and j the column index. Moving along the kth row of the matrix E of cross-efficiencies in Table 1 each entry E_{kj} is the productivity that results for unit j using the optimal weights of unit *k* accords to DMU *j*. The weights for the row unit are used for all calculations. Averaging along a row *k* is DMU *k*'s averaged appraisal of peers, termed EROW (*k*) in Sexton et al., averaging down column *j* is an averaged appraisal of DMU *j* of all row units, termed ECOL (*j*) in Sexton et al., i.e. peers' appraisal of unit *j*, weights for the row unit is used down the column and the cross efficiencies in a column are termed *peer appraisal* in Doyle and Green (1994). The standard Farrell efficiency score (h_{j0} in the original Charnes et al. 1978) formulation) is here $\boldsymbol{E_{kk}}$, i.e. the leading diagonal of the matrix (in fat types) gives us all efficiency scores of the *n* units, and are called *self-appraisal*.

Table 1. The cross-efficiency matrix E

		Peers' appraisal of column units				
	Unit	1	2	….*j*….	*n*	Average
Row units' appraisal of peers	1 2 . *k* . *n*	$\boldsymbol{E_{11}}$ E_{21} .. .. .. E_{n1}	E_{12} $\boldsymbol{E_{22}}$ .. .. .. E_{n2}	……… …….. …….. …$\boldsymbol{E_{kj}}$… …….. ……..	E_{1n} E_{2n} .. .. .. $\boldsymbol{E_{nn}}$	$\text{EROW}(k) = \frac{1}{n}\sum_j E_{kj}$
	Average	$\text{ECOL}(j) = \frac{1}{n}\sum_k E_{kj}$				EBAR

Source: Sexton et al. (1986), Doyle and Green (1994)

The unit *k*'s cross-efficiency score is defined as the average of the scores along the row of unit *k*:

$$\text{EROW}(k) = \frac{1}{n}\sum_{j=1}^{n} E_{kj} = \frac{1}{n}\frac{\sum_{j=1}^{n}\sum_{r=1}^{s} u_{rk}^{*} y_{rj}}{\sum_{j=1}^{n}\sum_{i=1}^{m} v_{ik}^{*} x_{ij}},\ k = 1,\dots,n \tag{30}$$

where the superscript '*' indicates the optimal solutions for the weights of unit *k*.[15] The peers' appraisal cross-efficiency score for unit *j* is the average for its column:

$$\text{ECOL}(j) = \frac{1}{n}\sum_{k=1}^{n} E_{kj} = \frac{1}{n}\frac{\sum_{k=1}^{n}\sum_{r=1}^{s} u_{rk}^{*} y_{rk}}{\sum_{k=1}^{n}\sum_{i=1}^{m} v_{ik}^{*} x_{ik}},\ j = 1,\dots,n \tag{31}$$

In Doyle and Green (1994) the efficiency scores for unit *k*, respectively *j*, E_{kk} and E_{jj}, are skipped, but not in Sexton et al (1986). However, it is not always clear in the literature what

[15] In Doyle and Green the symbols used for output- and input weights are switched compared with the use in CCR and Sexton et al.

type of average to use in order to distinguish among the 100% efficient DMUs, average of one unit's appraisal of peers (row average) or peers' appraisal of one unit (column average). Doyle and Green use the row average while e.g. Wang and Chin (2010) show the row average in Table 1, p. 3668. Sexton et al. call the average of all the cross-efficiency values for EBAR.

In order to understand the nature of the weights the linear programming (LP) problems corresponding to the CCR model in (5) are set out. The production function exhibits constant returns to scale (CRS) and is input-oriented[16]. The primal problem using k as the unit under investigation is:

$$\begin{aligned} &\text{Min } \theta_k \\ &\text{subject to} \\ &\sum_{j=1}^{n} x_{ij}\lambda_j \leq \theta_k x_{ik}, i = 1,\dots,m \\ &\sum_{j=1}^{n} y_{rj}\lambda_j \geq y_{rk}, r = 1,\dots,s \\ &\lambda_j \geq 0 , \theta_k \text{ sign free} \end{aligned} \tag{32}$$

The solution for the endogenous variables λ_j and θ_k are functions of the exogenous variables, i.e., the observations of inputs and outputs for all units. For a specific observation k the left-hand sides in the two first types of constraints defines the reference point on the frontier for unit k. If unit k is inefficient the unit index of the positive intensity weights λ_j in the solution of problem (5) show the referencing units for the projection of observation k to a frontier face (a face of full dimension $m+s$-1 is called a facet). The intensity weights are unique for observation k because the weight for k itself is zero, but on the right-hand sides both inputs and outputs of unit k appear.[17] Referencing units will per definition lie on the frontier with the efficiency score of 1, and the two constraints in (32) will hold with equality.

The dual LP model involving the weights appearing in (32) is:

[16] When the frontier function is CRS input-orientation and output-orientation yield identical efficiency scores and cross-efficiency terms.

[17] To make this point clearer the weights λ_j should have a subscript k, but this rather obvious indexing is usually supressed in the DEA literature.

$$\text{Max} \sum_{r=1}^{s} u_{rk} y_{rk}$$

subject to

$$\sum_{i=1}^{m} v_{ik} x_{ik} = 1 \tag{33}$$

$$\sum_{r=1}^{s} u_{rk} y_{rj} - \sum_{i=1}^{m} v_{ik} x_{ij} \leq 0, j = 1,..,k,..,n$$

$$v_{ik}, u_{rk} \geq 0$$

We may notice that the optimal efficiency score for unit *k* is calculated just by the weighted sum of outputs because the weighted sum of inputs is normalised to 1 by the nature of the optimisation problem (5) when forming the dual problem to (32). The second constraint in (33) is expressing the property that no unit can have productivity higher than one using the weights of the unit *k* under investigation when defining productivity. Looking at the primal formulation (32) this is equivalent to the restriction that the best performance of any unit is to be on the frontier. For inefficient units with inequalities in the constraints in (32) the corresponding shadow prices or weights will be zero.

A critique of the cross-efficiency approach

In Sexton et al. (1986) cross-efficiencies are introduced in order to address the issue of evaluating price efficiency in the case of no observations of prices. In Doyle and Green (1994) advantages of cross-efficiencies are listed. The two main advantages of cross efficiency is claimed to be i) a unique ranking of efficient units is achieved, ii) unrealistic weight schemes are eliminated. Let us investigate these two claims.

First of all it must be made clear that measuring the efficiency of production units is based on finding the relative distance in a specific sense between each unit and the common frontier function for all units. The classical definition of technical efficiency in Farrell (1957) is based on projecting each unit to the frontier assuming a common contraction factor for inputs or expansion factor for outputs. In order to find the frontier and the optimal contraction/expansion factors the role of cross efficiencies, appearing implicitly in the second constraint in the dual LP problem, is to make sure that no unit can be more efficient than a unit on the frontier. Once the frontier is determined the efficiency score for a unit is calculated without considering cross efficiencies. There seems to be no further role to be played by the cross efficiencies. The average value of cross efficiencies has no apparent economic meaning that can be used to rank efficient units.

If for some reason a ranking of efficient units is wanted, there are other approaches. Charnes et al. (1985) construct a measure that describes the units' importance as benchmarks by utilizing the fact that the efficient units act as referents for the inefficient units and simply count the number of times a unit acts as a referent. Andersen and Petersen (1993) calculate super-efficiency scores by excluding one efficient unit at a time from the set of observations used to calculate the frontier. Torgersen et al. (1996) calculate the importance of efficient units as benchmarks for inefficient units by calculating the improvement potential of the inefficient units that has the efficient unit under investigation in their reference set. (See Torgersen et al. for more references literature on how to discriminate between efficient units.)

To make the critique more precise let us look at the interpretation of a cell in the E – matrix; $E_{kj} = \sum_{r=1}^{s} u_{rk}^{*} y_{rj} / \sum_{i=1}^{m} v_{ik}^{*} x_{ij}$. Let us assume that unit j is efficient, i.e. $\boldsymbol{E_{jj}} = 1$. This will then be the highest value in the column, and the other cross-efficiency values will be created using the weights found optimal for the row units. The weights of unit k (row) are used to evaluate the productivity of unit j (column). But the weights are obtained from calculating the efficiency score θ_k for unit k, and they are shadow prices on output and input constraints in the optimisation problem (32) for unit k. Their interpretation is the change in the efficiency score of a marginal change in an input or output evaluated at a frontier point that is the projection point of the inefficient row unit in question (see Section 3 on the Interpretation of shadow prices). A higher value of E_{kj} than for other row units in the column for unit j must mean that one or more weights for outputs must be higher or one or more weights for inputs must be lower than for the other row units. But the weights in peer-appraisal cross-efficiencies going down a column all stem from different optimisation problems, and all have shadow-price interpretations that only have a meaning in these separate optimisation problems. It should also be pointed out that the cross-efficiency score E_{kj} is a productivity measure and not an efficiency measure because the denominator in (29) is typically not equal to one as it is for unit k. Thus, the column sum of cross efficiency scores does not have any apparent economic meaning. The projection points of the row units may typically be on different faces and different from the one or more faces the efficient unit j is located on. In the case of unit j being inefficient it is even less likely that its projection point belongs to the same face as all row units in the column for unit j. If the idea is that a cross-efficiency number should tell us how close the row unit is to the frontier compared with the column unit, this seems to be an idea without sound foundation. A cross-efficiency score tells nothing about the position of unit *k* relative to the relevant face on the frontier. Aggregation down the column makes this even worse.

It is stated the following in the web-site deazone.com about the cross-efficiency matrix:

> "This tool for interpreting the results consists of creating a table where the number of rows (j) [i][18] and columns (j) equals the number of units in the analysis. For each cell (ij), the efficiency of unit j is computed with weights that are optimal to unit j. The higher the values in a given column j, the more likely it is that the unit j is an example of truly efficient operating practices (Doyle and Green 1994)".

However, there is no theoretical substantiation of the claim in the last sentence, so the statement that to be the most cross-efficient DMU is a great mark of distinction, and harder to attain than 100% efficiency (Doyle and Green 1994) is just hanging in the air, as is the statement that cross efficiency provides an effective measure for differentiating performance among all DMUs (Du et al 2013).

As in the literature on weight restriction in DEA there is an underlying tendency to introduce preferences also in cross-efficiency. However, it seems rather meaningless to say that "cross efficiency takes into account the preferences (regarding multipliers) from all DMUs" based on the averaging procedure of cross-efficiencies (Du et al 2013, p.2).

Secondary goals

A problem for giving cross-efficiencies any meaning is that weights may not be unique (weights for units being vertex points will in general have non-unique weights). Therefore Sexton et al (1986) introduced secondary objectives based on finding the minimum or maximum values of the weights for a unit in order to influence cross-efficiency scores. The optimisation problems can be set up as in Wang and Chin (2010a, c):

[18] A row should be labelled i.

$$\begin{aligned}
&\text{Min} \sum_{r=1}^{s} (u_{rk} \sum_{j=1, j \neq k}^{n} y_{rj}) \\
&\text{subject to} \\
&\sum_{i=1}^{m} (v_{ik} \sum_{j=1, j \neq k}^{n} x_{ij}) = 1 \\
&\sum_{r=1}^{s} u_{rk} y_{rk} - \theta_{kk}^{*} \sum_{i=1}^{m} v_{ik} x_{ik} = 0 \\
&\sum_{r=1}^{s} u_{rk} y_{rj} - \sum_{i=1}^{m} v_{ik} x_{ij} \leq 0, \ j = 1, \ldots, n; j \neq k \\
&u_{rk} \geq 0, \ r = 1, \ldots, s \\
&v_{ik} \geq 0, \ s = 1, \ldots, m
\end{aligned} \tag{34}$$

This problem is in the spirit of the suggestion in Doyle and Green (1994) called Method III, with a normalisation to 1 of the denominator in an aggregate productivity expression. The efficiency level for the unit *k* under investigation is to be kept at the value of the efficiency score found as the solution to problem (32), θ_{kk}. The condition on all individual cross-efficiencies is the same as in (33).

Sexton et al. (1986) termed this choice of weights for an *aggressive* strategy of the units. However, it is of course the analyst that performs this choice of weights making the cross-efficiencies as small as possible under the constraints.

The opposite option of making the cross-efficiencies as large as possible was termed *benevolent*:

$$\begin{aligned}
&\text{Max} \sum_{r=1}^{s} (u_{rk} \sum_{j=1, j \neq k}^{n} y_{rj}) \\
&\text{subject to} \\
&\sum_{i=1}^{m} (v_{ik} \sum_{j=1, j \neq k}^{n} x_{ij}) = 1 \\
&\sum_{r=1}^{s} u_{rk} y_{rk} - \theta_{kk}^{*} \sum_{i=1}^{m} v_{ik} x_{ik} = 0 \\
&\sum_{r=1}^{s} u_{rk} y_{rj} - \sum_{i=1}^{m} v_{ik} x_{ij} \leq 0, \ j = 1, \ldots, n; j \neq k \\
&u_{rk} \geq 0, \ r = 1, \ldots, s \\
&v_{ik} \geq 0, \ s = 1, \ldots, m
\end{aligned} \tag{35}$$

However, although running these models overcome the non-uniqueness of the weights, the approach does not contribute to understand why the cross-efficiencies have any good meaning in the first place. It is also not clear that the weights u_{rk} and v_{ik} that are endogenous weights in problems (7) and (8) realise the condition $\sum_{i=1}^{m} v_{ik} x_{ik} = 1$.

Liang et al (2008a) also discuss secondary goals using the sum of Farrell technical inefficiencies over the units as the objective function to be minimised. It is stated as a theorem that the approach is equivalent to a form of the Doyle and Green (1994) model. However, neither a proof nor information of which model of Doyle and Green they have in mind are shared with the reader. Wang and Chin (2010a) elaborate further on the approach of Liang et al.

Elimination of unrealistic weight schemes

The cross-efficiency score has been proposed to be used as the ranking criteria usually including the standard efficiency score. Then "unbalanced" weighting schemes may not have a significant impact on the ranking. Doyle and Green (1994) are especially concerned with *maverick* DMUs "choosing to use zero weights for variables that will set them in a bad light". Using only cross-efficiency scores will allegedly avoid ranking such units so favourably. But this is an empirical question and is not shown to hold in general.

However, it may be the case that the occurrence of too many zero weights may be the consequence of ill-conditioned data (Olesen and Petersen 1996) and not something that can be easily fixed by deviating from the narrow path of efficiency evaluation proper.

Using cross-efficiency profiles

Doyle and Green (1994) suggested that additional characteristics of cross-efficiencies like variance and range of ratings may be considered. Such criteria were also mentioned in Despotis (2002). The approach of Ramón et al. (2010) tries to prevent unrealistic weighting schemes by focussing on the choice of profiles of the weights. This is achieved by introducing weight restrictions on the multipliers. Although the purpose of the restrictions is to satisfy an idea about a profile the critique in Doyle and Green (1994) that any weight restriction scheme is arbitrary still holds.

Variable returns to scale

The constant returns to scale model of Charnes et al. (1978) is referred to in most of the papers on cross-efficiency. But the approach is extended to a variable returns to scale (VRS) model of Banker et al. (1984) in Wu et al. (2009a) doing cross-efficiency on Olympic rankings. The dual model (6) is made into a VRS model by introducing a shadow price (weight) u_k on the convexity constraint in the primal model (5) of the sum of intensity weights being 1, resulting in this variable to be subtracted from the objective function in (6) and from the second constraint. Cross-efficiencies are now formed by

$$\frac{\sum_{r=1}^{s} u_{rk} y_{rj} - u_k}{\sum_{i=1}^{m} v_{ik} x_{ij}}, \; k, j = 1,..,n \tag{36}$$

This is called the VRS aggregate output to input ratio in Wu et al (2010a). But because this shadow price is free in sign, the numerator may, for some combination of *k*s and *j*s, become negative. This is then taken care of in the model calculating the weights by introducing a constraint that the numerator should be non-negative. This works technically but implies that the production possibility set is no longer the same as the set for a VRS specification of the DEA model. This adds to the critique of the cross-efficiency approach.

Game applications

Many of the cross-efficiency papers are written as if the unit or DMU itself considers setting the weights. Maybe this is the reason for game theory being introduced in the cross-efficiency models. Construction of game models can be found in Liang et al. (2008b), Wang and Chin (2010b), and Wu et al. (2009 a, b, c). In these games each DMU is regarded as being in competition with each other, and setting its bundle of weights that optimises its own efficiency score under various constraints concerning efficiency scores of the other units. In Wu et al. (2009c) the cross-efficiency score for a unit is calculated including its standard efficiency score, and each DMU participates in a bargaining game about setting its efficiency score. The game has a Nash bargaining solution implying that in equilibrium all units accept a score that is in between the standard efficiency score and the cross-efficiency score.

Wu et al. (2009c) want to eliminate the average assumption in determining the ultimate cross-efficiency and to improve the cross efficiency evaluation method from a cooperative game perspective. Each DMU will be a player, the characteristic function value of each coalition is defined, and the solution of Shapely value is computed to determine the ultimate cross efficiency for each DMU.

When modelling games about determining efficiency scores one is certainly outside the frame of trying to evaluate the efficiency of a unit relative to a frontier production function.

Conclusions

The main purpose of introducing cross efficiency was to get a better discrimination between efficient units using a constant-returns-to-scale specification. The cross-efficiency terms appear in a condition in the dual LP problem securing that no unit can have productivity greater than 1 (a normalisation), calculating the productivities using the optimal weights of the unit under investigation. The weights are the shadow prices on output- and input constraints in the primal (enveloping) LP problem, and are zero for all non-binding constraints. The Farrell technical efficiency score for a unit is calculated as the common scalar contracting all the inputs proportionally in the case of input orientation, and the inverse of the common scalar expanding the outputs in the case of output orientation. In the cross-efficiency literature there has been no formal analysis supporting the view that the cross-efficiency numbers have any role to play in evaluating the efficiency of the units. The cross-efficiency terms will depend on both the levels and mix of the outputs and inputs of the units in question. It may be that when one starts using the ratio definition (1) of efficiency the notion of productivities unduly capture the imagination of the analyst. In the literature using cross-efficiency the units are often given an active role of choosing the most favourable weights and choosing weights in order to put other units in an unfavourable or favourable light. Of course, it is the analyst that is performing an efficiency analysis, and weights are the result of the specific method used, not any choice of units to make. When various types of games involving units as competing players setting weights are introduced such games has very little to do with efficiency measurement proper.

7. Conclusions

Economists' approach to efficiency analysis is commonly based on starting with a transformation of inputs into outputs that is called the transformation function. This concept is to be understood in a broad context applicable to any type transformation process run by humans. Farrell (1957) is then the natural reference to efficiency concepts connected with the notion of a frontier production function introduced by Farrell. In a linear programming (LP) context this means that the envelopment form is the primal problem. The ratio form being the set-up in Charnes et al. (1978) then represents the dual approach. However, it is argued in Section 2 that the ratio form as formulated originally is not the way to estimate an efficiency score. Therefore the ratio approach is regarded as a detour and the analyst should stick to formulating the estimation problem as a LP problem. Standard software usually gives solutions to both problems.

For the purpose of interpretation it is useful to remember the definition of shadow prices in LP problems. As elaborated upon in Section 3 the shadow prices on inputs and outputs show the change in the efficiency score of a change in the given inputs and outputs in the respective constraints.

The ratio form of Charnes et al. (1978) shown in Section 2 brought into the efficiency story the concept of productivity. The connection is that efficiency can be defined as the ratio of observed productivity and the productivity of the projection of an inefficient observation to the frontier function either keeping outputs fixed (input orientation) or the inputs fixed (output orientation). This connection is elaborated upon in Section 4 showing the relative productivity interpretation for the generalised five Farrell efficiency measures covering the case of variable returns to scale. We then have technical efficiency measures, scale efficiency measures and a technical measure of productivity, the last two types of measures building upon the old concept of technically optimal scale in production theory.

Two subfields of efficiency analysis are surveyed in Section 5 and 6 and found to have serious shortcomings. In Section 5 the introduction of so called "weight restrictions" are surveyed and the shortcomings thoroughly discussed. The idea of weight restriction stems from the ratio form of the efficiency problem and the common occurrence of optimal solutions of zero for the weights or dual variables or shadow prices. The consequence of imposing constraints on dual variables is that if the constraints are binding, then it is the analyst that imposes his choice of values on the solution. But from an engineering point of view it is the input-output space that is the reality, and if the more or less complex connection between the imposed values of dual variables and the production possibility set is not known, then using weight restrictions may lead to nonsensical results.

The meaning of the concept of cross efficiency is discussed in Section 6. This concept also has its origin in following the ratio formulation of the estimation of efficiency. The basic idea is that something can be learned by using the weights of inputs and outputs found as the optimal solution for a unit to evaluate the other units, thus forming a matrix of cross efficiencies. However, this is a misuse of the efficiency analysis proper. The weights are unit specific in the problem of determining each unit's efficiency score. It is very difficult to see that cross efficiency calculations can provide any insights into the purpose of estimation efficiency scores.

References

Andersen P and Petersen NC (1993) A procedure for ranking efficient units in Data Envelopment Analysis. *Management Science* 39, 1261-1264.

Allen R, Athanassopoulos A, Dyson RG and Thanassoulis E (1997) Weight restrictions and value judgments in data envelopment analysis: evolution, development and future directions. *Annals of Operations Research* 73, 13-34

Banker RD, Charnes A, and Cooper WW (1984) Some models for estimating technical and scale inefficiency in data envelopment analysis. *Management Science* 30 (9), 1078-1092

Beasley JE (1995) Determining teaching and research efficiencies. *Journal of the Operational Research Society* 46, 441-452

Boles JN (1967) Efficiency squared—efficient computation of efficiency indexes. *Western Farm Economic Association, Proceedings 1966*, 137-142

Boles JN (1971) *The 1130 Farrell efficiency system – multiple products, multiple factors*. Giannini Foundation of Agricultural Economics, February 1971

Charnes A, Cooper WW, and Rhodes E (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research* 2, 429-444

Charnes A, Cooper WW, and Rhodes E (1979) Short communication. Measuring the efficiency of decision making units. *European Journal of Operational Research* 3, 339

Charnes A, Cooper WW, Huang ZM and Sun DB (1990) Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks. *Journal of Econometrics* 46, 73–91

Charnes A, Clark CT, Cooper WW and Golany B (1985) A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the U.S. Air Forces. *Annals of Operations Research* 2, 95-112.

Chen T-Y (2002) An assessment of technical efficiency and cross-efficiency in Taiwan's electricity distribution sector. *European Journal of Operational Research*, 137, 421–433.

Cooper WW, Seiford LM, Tone K (2006) *Introduction to data envelopment analysis and its uses. With DEA-solver software and references*. Springer Science + Business Media, New York

Despotis DK (2002) Improving the discriminating power of DEA: focus on globally efficient units. *Journal of the Operational Research Society* 53(2), 314-323.

Doyle J and Green R (1994) Efficiency and cross-efficiency in DEA: derivations, meanings and uses. *Journal of the Operational Research Society* 45(5), 567-578.

Du J, Cook WD, Liang L and Zhu J (2014) Fixed cost and resource allocation based on DEA cross-efficiency. *European Journal of Operational Research* 235(1), 206-214

Dyson RG and Thanassoulis E (1988) Reducing weight flexibility in data envelopment analysis. *Journal of the Operational Research Society* 39 (6), 563-576

Färe R, Grosskopf S, Noh D-W and Yaisawarng S (1993) Derivation of shadow prices for undesirable outputs: a distance function approach. *Review of Economics and Statistics* 75(2), 374-380

Farrell MJ (1957) The measurement of productive efficiency. *Journal of the Royal Statistical Society*, *Series A (General)* 120 (III), 253-281(290)

Farrell MJ and Fieldhouse M (1962) Estimating efficient production functions under increasing returns to scale. *Journal of the Royal Statistical Society*, *Series A (General)* 125 (2), 252-267

Frisch R (1965) *Theory of Production*. Dordrecht: D. Reidel

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

Førsund FR (2010) Dynamic efficiency measurement. *Indian Economic Review* 45(2), 125-159

Førsund FR (2013) Weight restrictions in DEA: misplaced emphasis? *Journal of Productivity Analysis* 40(3), 271-283

Førsund FR and Hjalmarsson L (1974) On the measurement of productive efficiency. *Swedish Journal of Economics* 76, (2), 141-154

Førsund FR and Hjalmarsson L (1979) Generalized Farrell measures of efficiency: an application to milk processing in Swedish dairy plants. *Economic Journal* 89, 294-315.

Førsund FR and Hjalmarsson L (2004) Are all scales optimal in DEA? Theory and empirical evidence. Journal of Productivity Analysis, 21, 25–48

Førsund FR and Sarafoglou N (2002) On the origins of Data Envelopment Analysis. *Journal of Productivity Analysis* 17, 23-40

Førsund FR and Sarafoglou N (2005) The tale of two research communities: the diffusion of research on productive efficiency. *International Journal of Production Economics* 98(1), 17-40

Førsund FR, Kittelsen SAC and Krivonozhko VE (2009) Farrell revisited−visualising properties of DEA production frontiers. *Journal of the Operational Research Society* 60, 1535-1545

Førsund FR, Hjalmarsson L, Krivonozhko VE, Utkin OB (2007) Calculation of scale elasticities in DEA models: direct and indirect approaches. *Journal of Productivity Analysis* 28, 45–56

Hanoch G (1970) Homotheticity in joint production. *Journal of Economic Theory* 2, 423–426.

Joro T and Viitala E-J (2004) Weight-restricted DEA in action: from expert opinions to mathematical models. *Journal of the Operational Research Society* 55, 814-821

Krivonozhko V, Volodin AV, Sablin IA, Patrin M (2004) Constructions of economic functions and calculation of marginal rates in DEA using parametric optimization methods. *Journal of the Operation Research Society* 55(10),1049–1058

Liang L, Wu J., Cook WD and Zhu J (2008a) Alternative secondary goals in DEA cross-efficiency evaluation. *International Journal of Production Economics*, 113, 1025–1030

Liang L, Wu J, Cook WD and Zhu J (2008b) The DEA game cross-efficiency model and its Nash equilibrium. *Operations Research* 56(5), 1278–1288

McFadden D (1978) Cost, revenue and profit functions. In M. Fuss and D. McFadden (eds), *Production economics: A dual approach to theory and applications*, Vol. 1, Chapter 1. Amsterdam: North Holland Publishing Company, 3–109

Olesen OB and Petersen NC (1996) Indicators of ill-conditioned data sets and model misspecification in data envelopment analysis: an extended facet approach. *Management Science* 42(2), 205-219

Panzar JC and Willig RD (1977) Economies of scale in multi-output production. *Quarterly Journal of Economics* XLI, 481-493

Pedraja-Chaparro F, Salinas-Jimenez J and Smith P (1997) On the role of weight restrictions in data envelopment analysis. *Journal of Productivity Analysis* 8, 215-230

Podinovski VV (2004a) Production trade-offs and weight restrictions in data envelopment analysis. *Journal of the Operational Research Society* 55, 1311-1322

Podinovski VV (2004b) Suitability and redundancy of non-homogeneous weight restrictions for measuring the relative efficiency in DEA. *European Journal of Operational Research* 154, 380-395

Podinovski VV (2005) The explicit role of weight bounds in models of data envelopment analysis. *Journal of the Operational Research Society* 56, 1408-1418

Podinovski VV (2007) Improving data envelopment analysis by the use of production trade-offs. *Journal of the Operational Research Society* 58, 1261-1270

Podinovski VV and Athanassopoulos AD (1998) Assessing the relative efficiency of decision making units using DEA models with weight restrictions. *Journal of the Operational Research Society* 49, 500-508

Podinovski VV and Førsund FR (2010) Differential characteristics of efficient frontiers in data envelopment analysis. *Operations Research* 58(6), 1743–1754

Podinovski VV, Førsund FR and Krivonozhko VE (2009) A simple derivation of scale elasticity in data envelopment analysis. *European Journal of Operational Research* 197, 149–153

Ramón N, Ruiz JL and Sirvent I (2010) On the choice of weights profiles in cross-efficiency evaluations. *European Journal of Operational Research* 207 (3), 1564–1572

Sexton TR, Silkman RH and Hogan AJ (1986). Data envelopment analysis: critique and extensions *New Directions for Program Evaluation,* Special Issue: Measuring Efficiency: An Assessment of Data Envelopment Analysis 32 (Winter), 73-105

Shephard RW (1970) *Theory of cost and production functions* (first edition 1953). Princeton University Press: New Jersey

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

Sun S and Lu WM (2005) A cross-efficiency profiling for increasing discrimination in data envelopment analysis. *INFOR* 43(1), 51–60

Thanassoulis E, Portela MC and Allen R (2004) Incorporating value judgments in DEA. In: Cooper WW, Seiford LM, Zhu J (eds) *Handbook on data envelopment analysis*, Chapter 4, 99-138. Kluwer Academic Publishers, Boston/Dordrecht/London

Thanassoulis E, Portela MCS and Despić O (2008). Data envelopment analysis: the mathematical programming approach to efficiency analysis. In Fried HO, Lovell CAK, Schmidt SS (eds) *The measurement of productive efficiency and productivity growth,* Section 3.7, 321-340. Oxford University Press, Oxford

Thompson RG, Singleton Jr FR, Thrall RM and Smith BA (1986) Comparative site evaluation for locating a high-energy physics lab in Texas. *Interfaces* 16(6), 35-49

Torgersen AM, Førsund FR and Kittelsen SAC (1996) Slack-adjusted efficiency measures and ranking of efficient units. *Journal of Productivity Analysis* 7, 379-398

Wang Y-M and Chin K-S (2010a) Some alternative models for DEA cross-efficiency evaluation. *International Journal of Production Economics*, 128, 332-338

Wang Y-M and Chin K-S (2010b) Determination of the weights for the ultimate cross efficiency using Shapley value in cooperative game. *Expert Systems with Applications* 36(1), 872–876

Wang Y-M and Chin K-S (2010c) A neutral DEA model for cross-efficiency evaluation and its extension. *Expert Systems with Applications* 37 (5) 3666–3675

Wong YHB and Beasley JE (1990) Restricting weight flexibility in data envelopment analysis. *Journal of the Operational Research Society* 41, 829-835

Wu J, Liang L, and Chen Y (2009a) DEA game cross-efficiency approach to Olympic rankings. *Omega* 37(4), 909-918

Wu J, Liang L, Yang F (2009b) Determination of the weights for the ultimate cross efficiency using Shapley value in cooperative game. *Expert Systems with Applications* 36 (1), 872–876

Wu J, Liang L, Yang F and Yan H (2009c) Bargaining game model in the evaluation of decision making units. *Expert Systems with Applications*, 36, 4357–4362

Zhou P, Zhou X and Fan LW 2014. On estimating shadow prices of undesirable outputs with efficiency models: a literature review. *Applied Energy* 130, 799-806