MEMORANDUM

No 05/2011

Weight Restrictions in DEA: Misplaced Emphasis?

Finn R. 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 04/11	Vladimir Krivonozhko, Finn R. Førsund, Andrey V. Lychev *Terminal units in DEA: Definition and Determination*
No 03/11	André K. Anundsen, Tord S. H. Krogh, Ragnar Nymoen, Jon Vislie *Overdeterminacy and endogenous cycles: Trygve Haavelmo's business cycle model and its implications for monetary policy*
No 02/11	Nils Chr. Framstad *Portfolio Separation Properties of the Skew-Elliptical Distributions*
No 01/11	Karine Nyborg , Tao Zhang *Is corporate social responsibility associated with lower wages?*
No 22/10	Rebecca Graziani, Nico Keilman *The sensitivity of the Scaled Model of Error with respect to the choice of the correlation parameters: A Simulation Study*
No 21/10	Jennifer L. Castle, Jurgen A. Doornik , David F. Hendry, Ragnar Nymoen *Testing the Invariance of Expectations Models of Inflation*
No 20/10	Erik Biørn *Identifying Trend and Age Effects in Sickness Absence from Individual Data: Some Econometric Problems*
No 19/10	Michael Hoel, Svenn Jensen *Cutting Costs of Catching Carbon* *Intertemporal effects under imperfect climate policy*
No 18/10	Hans Jarle Kind, Tore Nilssen, Lars Sørgard *Price Coordination in Two-Sided Markets: Competition in the TV Industry*
No 17/10	Vladimir Krivonozhko, Finn R. Førsund and Andrey V. Lychev *A Note on Imposing Strong Complementary Slackness Conditions in DEA*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/forskning/publikasjoner/memorandum

WEIGHT RESTRICTIONS IN DEA:

MISPLACED EMPHASIS? *

by

Finn R Førsund

Department of Economics, University of Oslo

February 2011

Abstract: Measuring productive efficiency is an important research strand within fields of economics, management science and operations research. One definition of efficiency is the proportional scaling needed for observations of an inefficient unit to be projected onto an efficient production function and another definition is a ratio index of weighted outputs on weighted inputs. When linear programming is used to estimate efficiency the two definitions give identical results due to the fundamental duality of linear programming. Empirical applications of DEA using linear programming showed a prevalence of zero weights leading to questioning the consequence for the efficiency score estimate based on the ratio definition. Early literature on weight restrictions is exclusively based on the ratio efficiency. It was stated that variables with zero weights had no influence on the efficiency score, in spite of the alleged importance of the variables. This has been one motivation for introducing restrictions on weights. Another empirical result was that often there were too many efficient units. This problem could also be overcome by introducing weight restrictions. Weight restrictions were said to introduce values for inputs and outputs. The paper makes a critical examinations of these claims based on defining efficiency relative to a frontier production function.

Keywords: Weight restrictions; DEA; efficiency; frontier production function; primal and dual linear programming problems

JEL classification: C61, D20

* The paper is written within the research programme “Kostnadseffektiv drift av Forsvaret (Koster III)” (Cost Efficiency in Defence) at the Norwegian Defence Research Establishment (FFI).

1. Introduction

Measuring productive efficiency is an important research strand within the fields of economics, management science and operations research. The two seminal contributions are Farrell (1957) and Charnes et al. (1978). However, the fundamental definition of efficiency is apparently different in the two 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 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.

However, it is well known that if the proportional scaling factors can be estimated using linear programming, then if optimal solutions exist, the two definitions give identical efficiency scores due to the fundamental duality between the primal and dual problem of linear programming. It may then be said that is does not matter which definition that is adopted. But it will be argued in the paper that the choice of a definition does matter for the understanding of the problem at hand, because adopting the index definition and searching for imposing value judgements about relative values of inputs and outputs by means of various forms of weight restrictions lead to problem formulations that are not well founded, and may be regarded in conflict with principles of estimation of efficiency based on a data set.

The ratio expression defining efficiency in Charnes et al. (1978) is what is termed productivity in economics. The construction of productivity indices is 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. This is just what the constraints on the productivity index imply in the Charnes et al. (1978) definition. There are two basic approaches to this construction: estimating weights used to aggregate outputs and inputs, respectively, directly, or starting from an estimate of the relevant production function. Using output and input prices as weights is the standard example of the first approach, while utilising the Farrell technical efficiency measures as done when calculating the Malmquist productivity index (Caves et al. 1982) is an example of the latter.

The Farrell (1957) definition of efficiency is independent of any index of efficiency constructed directly by weighted outputs over weighted inputs, i.e., if an estimate of a frontier production function and input-output data for a unit are available, the Farrell technical efficiency scores can be calculated. The use of linear programming is just one of several tools of estimating a frontier function, but its use is not necessary for efficiency measurement in principle. The choice of estimation method depends on the specification of the form of the frontier production function, but the definition of efficiency remains the same.

Most of the papers concerned with weight restrictions start with the Charnes et al. (1978) efficiency definition of weighted outputs over weighted inputs. 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. It may help to think about public sector service-producing units that do not sell their outputs in markets. Two kinds of efficiency problems can then be posed. The first one concerns the efficiency of producing the services, i.e., the efficiency of utilising inputs in the production of the actual services provided, and the second problem addresses whether the set of services produced, promote the fundamental objectives of undertaking the service production in the first place. This is a question of *effectiveness*. The relative value of the services can only be found by connecting to the fundamental objectives of undertaking service production. However, almost all empirical applications of efficiency studies are concerned with the first type of problem; the efficient use of inputs in the production of the actual services provided. Then the weights should not be manipulated within this problem set up for yielding measures of technical efficiency, but be reserved for a second stage analysis.[1] In the case studied in Farrell (1957) there is one output, and known input prices. Farrell goes on to develop the second-stage analysis by estimating *allocative efficiency* of the employment of inputs, and finally multiplying technical and

[1] In Joro and Viitala (2004, p. 814) weight restrictions based on preferences or value information, is, somewhat at odds with the philosophy presented above, being regarded there as an “intermediate approach” between standard DEA analysis of technical efficiency and the analysis of economic efficiency said to requiring price information on inputs and outputs.

allocative efficiency to yield *overall efficiency* (see Diagram 1 in Farrell 1957, p. 254). (It is straightforward to develop a similar overall measure to cover optimal mix of multiple outputs priced in markets.)

The lesson to be learned for technical efficiency analyses is that the extension to an overall efficiency measure incorporating values does not interfere at all with the analysis of technical efficiency. It should also be noted that an overall optimal solution covering both stages requires that the provision of the service mix actually produced obeys technical efficiency.

Concerning estimation of efficiency it is stated in Charnes et al. (1990, p.73) that there are two developments to consider: empirical production functions and managerial performance evaluation (efficiency analysis). Thus, it is not the situation that the need to estimate a frontier technology is not well known in the literature. However, most of the papers are preoccupied with formulating constraints on weights in the weight space, and do not link this to implications for the production possibility set in input-output space, although it is realised in general that constraining the weights in the weight space implies expansion of the production possibility set in input-output space. But in Podinovski (2004a) (extended in Podinovski 2005) it is shown that starting with incorporating more information about the production possibilities in the form of trade-offs between outputs or inputs the corresponding weight restrictions can be found. This is called the trade-off approach. Podinovski (2004a, p. 1316) is very clear on the difference between what he calls 'technology thinking' and 'value thinking'. He, furthermore, warns that the latter approach may lead to inconsistencies with the production realities and distort the economic meaning of efficiency.

An alleged problem with the ratio definition of efficiency is that zero values of weights may often appear as optimal solutions. It is, then, important to have in mind that the data at hand actually determine the outcome (Olesen and Petersen 1996). Zero weights appear because data do not contain sufficient information to avoid this. 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.

The purpose of the paper is to have a closer discussion of the differences between the two seminal contributions to efficiency definitions, and in particular to address some aspects of the weight-restriction literature that seem to be misunderstandings of the issues involved. In Section 2 the efficiency definitions in the two seminal papers are reviewed, and in Section 3 the interpretation of the Farrell (1957) definitions in the case of a non-parametric benchmark technology are treated in detail in order to see the connection with the exposition in Charnes et al. (1978), thereby preparing the ground for exposing misunderstandings in the DEA weight-restriction literature. These misunderstandings are then commented upon in Section 4, and Section 5 concludes.

2. The seminal contributions to measurement of productive efficiency

The frame of reference for defining productive efficiency in Farrell (1957) is the most efficient production technology, now commonly termed the frontier production technology. Knowledge about such a technology can be obtained in two ways; based on engineering information about the production activity in question, or based on observed best practice. Given that an efficient technology is known, then efficiency of a unit is defined by the scaling that is required to project an inefficient observation to the frontier, either by a proportional scaling down of all inputs for given output levels by a scalar scaling factor, or by a proportional scaling up by a scalar scaling factor of all the outputs given the input quantities. Restricting the efficiency measures to be between zero and one the latter output-oriented efficiency measure is then the inverse of the scaling factor. These efficiency-measure definitions are independent of how the efficient technology, or frontier technology, is estimated, and, furthermore, independent of the scale properties of the frontier function. Farrell (1957) termed the measures for technical efficiency measures. Obviously, there is no value judgement involved about inputs or outputs.

The Farrell technical efficiency measures can in a general setting be formalised as follows:[2]

$$
\begin{aligned}
E_{1j} &= \mathrm{Min}_{\theta_j} \left\{ \theta_j : F_j(Y_j, \theta_j X_j) \le 0 \right\} \\
E_{2j} &= 1/\,\mathrm{Max}_{\phi_j} \left\{ \phi_j : F_j(\phi_j Y_j, X_j) \le 0 \right\},\ j = (1,\ldots,n)
\end{aligned} \qquad (1)
$$

Here $F_j(Y_j, X_j) = 0$ is the efficient multi-output, multi-input production function (with standard properties) for unit j , $F_j(Y_j, X_j) < 0$ indicates inefficient operations, Y_j and X_j are the output and input vectors respectively, the input-oriented technical efficiency measure is E_{1j}, the output-oriented efficiency measure is E_{2j}, and the scaling factors are θ_j and ϕ_j. The Farrell efficiency measures correspond to the concept of distance functions introduced in Shephard (1970).

In the case of input prices being available, Farrell (1957) also defines price or allocative efficiency based on cost minimisation relative to the frontier technology, and overall efficiency as the product of technical efficiency and price efficiency. Farrell (1957, pp. 260-261) puts forward arguments for weaknesses with 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.

Farrell (1957) also proposed a method of estimating a best-practice frontier by enveloping the data by a non-parametric piecewise function, imposing convex negatively-sloped isoquants, and constant returns to scale. He comments that this way of estimating a production function may not

[2] Farrell (1957) did not present formal definitions of the efficiency measures, may be 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 two widely reproduced graphical illustrations convey very well his efficiency definitions and his approach of estimation, specifying a piecewise linear convex unit isuquant enveloping the data points.

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).[3]

Farrell (1957) applied his approach in the case of only a single output. His method of estimating the efficiency scores was based on solving a system of linear equations. In the discussion of Farrell’s paper at the meeting of the Royal Statistical Association in 1957 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).

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 Encyclopedia 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

[3] 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.

ratio definition is only valid for constant returns to scale. The formal definition of efficiency for a unit j_0 is:[4]

$$\begin{aligned}
&\text{Max}\, h_{j_0} = \frac{\sum_{r=1}^{s} u_{rj_0} y_{rj_0}}{\sum_{i=1}^{m} v_{ij_0} x_{ij_0}} \\
&\text{subject to} \\
&\frac{\sum_{r=1}^{s} u_{rj_0} y_{rj}}{\sum_{i=1}^{m} v_{ij_0} x_{ij}} \leq 1 \;\; j = 1\ldots, j_0, \ldots n \\
&u_{rj_0}, v_{ij_0} \geq 0 \;\; \forall r, i
\end{aligned} \qquad (2)$$

In (2) 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]

As the definition stands this approach seems to be quite different from the seminal approach of Farrell (1957). Furthermore, Farrell makes quite strong negative comments about the idea of forming an efficiency index by "weighing together input factors." He sees the difficulty as finding a suitable set of prices serving as weights, and if they are found, he states that the whole exercise then boils down to making cost comparisons (Farrell 1957, p. 264).

But Charnes et al. (1978), as promised, manage to relate the analysis to economics by turning the fractional program of the ratio definition (2) into an equivalent linear program, 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

[4] In Podinovski (2001a) the first ratio in (2) is called 'absolute efficiency' (as stated earlier this ratio is defined as productivity in economics), while the full problem (2) is called relative efficiency. The interesting observation is made that problem (2) implies both a maximisation of absolute efficiency and relative efficiency.

[5] Later in the paper, in order to avoid the weights turning out zero, a non-Archimedean number ε is used as the lower limit for the weights in (2). However, since this number can be arbitrarily close to zero, for practical purposes this may leave economic rates at zero of 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).

efficiency scores of the different approaches are therefore identical due to the duality property of a linear programming problem.

However, the question still remains why Charnes et al. (1978) start with the so-called engineering approach and neglect Farrell (1957) completely concerning the basic definition of efficiency. The referred heat-ratio efficiency in engineering is quite obvious, but restricted to a single output and input, and the ratio definition (2) of efficiency does not really follow automatically. I do not know of any use of such a definition of efficiency in engineering. People following the ratio definition have not made any other references to engineering than the reference to Charnes et al. as far as I know, but still call the ratio definition for a classical engineering approach. In Banker et al. (1984) p. 1078, the following statement is made: "the CCR [Charnes et al. (1978)] ratio definition….generalizes the single-output to single input classical engineering-science ratio definition to multiple outputs and multiple inputs…" No reference to the engineering literature is offered.

The preoccupation with restrictions on weights in the operations research literature on DEA is obviously directly influenced by the ratio definition. This paper will try to demonstrate that starting with the weights may easily lead to misunderstandings about the relevant economic definition of efficiency and the estimation problems involved. Although the ratio definition is "saved" in a technical sense when linear programming is used to estimate the frontier production function and related technical efficiency measures, its status is in my opinion more in the "as if " category rather than being a primary definition of efficiency on its own. The usefulness of the dual relationships in linear programming is that characterisations of the frontier production function can be made in the form of marginal productivities, rates of substitution between pair of inputs, and rates of transformation between pair of outputs, as will be shown in Section 3.

An important observation is that the Farrell (1957) definition of efficiency does not depend on the method by which the frontier and efficiency scores are estimated, unlike starting from the ratio definition that can only be estimated when linear programming is used. It is also worthwhile to note that the ratio definition 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.

3. The interpretation of shadow prices

In order to expose the awkward understanding of the simultaneous problem of estimating a value-free frontier production function and technical Farrell efficiency scores in the weight-restriction literature, let us start with the production possibilities. We will assume a non-parametric frontier function as was Farrell's preferred alternative, and assume the axioms for the production possibility set presented in Banker et al. (1984) to hold. The production possibility set in the case of variable returns to scale can be written:

$$T = \left\{ (X,Y) : X \geq \sum_{j=1}^{n} X_j \lambda_j , Y \leq \sum_{j=1}^{n} Y_j \lambda_j , \sum_{j=1}^{n} \lambda_j = 1, \lambda_j \geq 0 \right\} \tag{3}$$

Here X_j is the input vector of m inputs, Y_j the output vector of s outputs, and $j = 1,\ldots, n$ is the index for the n units, using the notation in (1) and (2). 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 we impose constant returns to scale.

The standard primal problem of estimating Farrell technical efficiency scores for a unit j_0 in the case of variable returns to scale and input orientation of the efficiency measure, following the definition (1) of efficiency (input orientation), is:

$$
\begin{aligned}
&E_{1j_0} = \text{Min } \theta_{j_0} \\
&\text{subject to} \\
&\sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_{j_0} x_{ij_0} , i = 1,...,m \\
&\sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{rj_0} , r = 1,...,s \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j \ge 0 \text{ , } \theta_{j_0} \text{ sign free}
\end{aligned} \qquad (4)
$$

The variables ($x_{1j},\ldots, x_{mj}$) and ($y_{1j},\ldots, y_{sj}$) represent the observed m inputs and s outputs of production units $j = 1,\ldots,j_0,\ldots,n$. The endogenous intensity weights λ_j (telling us the composition of the frontier reference point for unit j_0) is unit-specific, but for notational ease the index for j_0 is suppressed, as is the usual practice in the literature.

The general idea of Farrell technical efficiency for a unit is based on measuring the relative distance between a unit and the benchmark frontier, and that the way of measuring is to scale down proportionately the inputs, keeping the output fixed, defining the input-oriented measure as θ_{j_0} in (4), and scaling up proportionally the outputs keeping the inputs fixed for the output-oriented measure. Thus, the Farrell definition of efficiency, based on a general frontier function as benchmark, is not defined by a weighted sum of outputs over a weighted sum of inputs.

However, in the case of estimating the frontier function by using linear programming, assuming that the frontier is piecewise linear, the fundamental duality theorem of linear programming of equality between the value of the objective function of the primal and the value of objective function of the dual, can be utilised to express the Farrell efficiency score formally as a weighted sum of outputs on a weighted sum of inputs.

The dual problem to problem (4) is[6]

[6] This is the equivalent linear programming problem to the fractional programming problem (2). It may be noted that the latter problem, as far as I 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 (2) is done just to state a definition of efficiency, but as computation is concerned it is a detour.

$$\begin{aligned}
&\text{Max } (\sum_{r=1}^{s} u_{rj_0} y_{rj_0} - u_{j_0}) \\
&\text{subject to} \\
&\sum_{i=1}^{m} v_{ij_0} x_{ij_0} = 1 \\
&\sum_{r=1}^{s} u_{rj_0} y_{rj} - \sum_{i=1}^{m} v_{ij_0} x_{ij} - u_{j_0} \leq 0, \ j = 1 \ldots j_0 \ldots n \\
&v_{ij_0}, u_{rj_0} \geq 0, \ u_{j_0} \text{ sign free}
\end{aligned} \tag{5}$$

The variables $v_{ij_0}, u_{rj_0}, u_{j_0}$ are the shadow prices on the constraints in (4), which are the input constraints, the output constraints and the convexity constraint, respectively. In DEA it is more common to call these variables for multipliers or weights, and in the paper these terms are used interchangeably. We then have the fundamental duality result for an 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 on the convexity constraint in the case of variable returns to scale, and this shadow price will be zero in the case of constant returns to scale.

The shadow prices on the two constraints in (5) will have the interpretation of the efficiency score for the first constraint and the intensity weight for the second constraint.

As to the concern of zero weights influencing the efficiency score it should be noted from (5) 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 is equal to one and the concern should only be about input weights.) In the DEA literature the product of a shadow price and an input (output) is called virtual input (output).

Setting up the Lagrangian for the constrained optimisation problem (4) for unit j_0 we have

$$\begin{aligned} L = & -\theta_{j_0} \\ & -\sum_{r=1}^{s} u_{rj_0} (y_{rj_0} - \sum_{j=1}^{n} \lambda_j y_{rj}) \\ & -\sum_{i=1}^{m} v_{ij_0} (\sum_{j=1}^{n} \lambda_j x_{ij} - \theta_{j_0} x_{ij_0}) \\ & -u_{j_0} (\sum_{j=1}^{n} \lambda_j - 1) \end{aligned} \tag{6}$$

The necessary first-order conditions for a solution to problem (4) are:

$$\begin{aligned} & \frac{\partial L}{\partial \lambda_j} = \sum_{r=1}^{s} u_{rj_0} y_{rj} - \sum_{i=1}^{m} v_{ij_0} x_{ij} - u_{j_0} \leq 0 \ (= 0 \text{ for } \lambda_j > 0),\ j = 1,..., j_0,..., n \\ & \frac{\partial L}{\partial \theta_{j_0}} = -1 + \sum_{i=1}^{m} v_{ij_0} x_{ij_0} = 0 \\ & u_{rj_0} \geq 0 \ (= 0 \text{ for } y_{rj_0} < \sum_{j=1}^{n} \lambda_j y_{rj}), r = 1,..., s \\ & v_{ij_0} \geq 0 \ (= 0 \text{ for } \sum_{j=1}^{n} \lambda_j x_{ij} < \theta_{j_0} x_{ij_0}), i = 1,..., m \end{aligned} \tag{7}$$

From the first condition we have that the intensity weight will be zero for unit j_0 if this unit is inefficient; using the duality result of equality of the two objective functions in (4) and (5) we have $\sum_r u_{rj_0} y_{rj_0} - u_{j_0} < \sum_i v_{ij_0} x_{ij_0}$. 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 slacks.

If we have a unique solution to problem (4) 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 (4), we get:

$$\begin{aligned}
&\frac{\partial(-\theta_{j_0})}{\partial y_{rj_0}} = \frac{\partial L}{\partial y_{rj_0}} = -u_{rj_0}(1-\lambda_{j_0}) = -u_{rj_0} \Rightarrow \frac{\partial \theta_{j_0}}{\partial y_{rj_0}} = u_{rj_0}\ , r=1,\ldots,s \\
&\frac{\partial(-\theta_{j_0})}{\partial x_{ij_0}} = \frac{\partial L}{\partial x_{ij_0}} = v_{ij_0}(\theta_{j_0} - \lambda_{j_0}) = v_{ij_0}\theta_{j_0} \Rightarrow \frac{\partial \theta_{j_0}}{\partial(\theta_{j_0} x_{ij_0})} = -v_{ij_0}\ , i=1,\ldots,m
\end{aligned} \tag{8}$$

We have that $\lambda_{j_0} = 0$ for unit j_0 being inefficient from the first necessary condition in (7). For inefficient units the shadow price on the output constraint is then directly interpreted as the increase in the efficiency score of a marginal increase in the output variable in question evaluated at a frontier point. The unit of measurement for the shadow price is efficiency score units (dimensionless) per measurement unit of the output variable in question.

Concerning a change in an input variable we have in general that the shadow price on an input constraint for an inefficient unit measures the impact on the efficiency score of a marginal increase in the input variable in question. In the direct interpretation the shadow price on the constraint is weighted with the efficiency score. But because this is 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 is again efficiency score units per measurement unit of the input variable in question.

As we see the interpretation of the shadow prices is quite clear, conforming to the standard interpretation of shadow prices in constrained optimisation problems. Inspecting the objective function of the dual (5) we have that the weighing of outputs transform the expression into units of the efficiency score, i.e., a dimensionless number. Weighing an output or input is just calculating the marginal contribution of the variable in question to the efficiency score at the optimum solution; this has nothing to do with any external economic value put on the variable. So-called virtual inputs (outputs) are just expressing the contribution to the efficiency score at the optimal solution of the variable in question.

The interpretation of the shadow prices in terms of standard production function concepts can straightforwardly be made utilising the dual problem (5). The second constraint will hold with equality in an optimal solution 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:

$$
\begin{aligned}
&\sum_{r=1}^{s} u_{rj_0} y_{rj} - \sum_{i=1}^{m} v_{ij_0} x_{ij} - u_{j_0} = 0 \\
&u_{rj_0} dy_{rj} - v_{ij_0} dx_{ij} = 0 \Rightarrow \frac{dy_{rj}}{dx_{ij}} = \frac{v_{ij_0}}{u_{rj_0}}
\end{aligned} \tag{9}
$$

This is the economic concept of the marginal productivity of input *i* in terms of the output of type *r*. Using also the types *r'* and *i'* of outputs and inputs, respectively, we develop in the same way the following expressions:

$$
\begin{aligned}
&u_{rj_0} dy_{rj} + u_{r'j_0} dy_{r'j} = 0 \Rightarrow \frac{dy_{rj}}{dy_{r'j}} = -\frac{u_{r'j_0}}{u_{rj_0}} \\
&-v_{ij_0} dx_{ij} - v_{i'j_0} dx_{i'j} = 0 \Rightarrow \frac{dx_{ij}}{dx_{i'j}} = \frac{v_{i'j_0}}{v_{ij_0}}
\end{aligned} \tag{10}
$$

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 facet (or face).

We know that the extreme-efficient units (efficient units with zero slacks) are vertex points (or located on ridges), 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.

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 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 lesson for efficiency definition is that following the ratio definition of Charnes et al. (1978), does not imply an estimation of weights independent of the production function. The weights provided by the dual solution to the envelopment problem has no independent value dimension, but have the standard interpretation of shadow prices on constraints in the envelopment problem, as is the terminology used in economics.

4. A critical examination of the literature

The first (published) paper to raise the weight-restriction issue in DEA is Thompson et al. (1986). This paper states a rather special version of the ratio efficiency definition in Charnes et al. (1978): "The explicit objective function in DEA is the ratio of present value of benefits to present value of costs" (Thompson et al. 1986, p. 43). The efficiency problem set up involves just six units, and three inputs are specified assuming the same output for all the units. Two of the inputs are not inputs in a production function sense, but represent different properties of sites for locating a high-energy physics lab in Texas. The variables are present value of construction costs, user time delay of completing research projects (measured as excess over time at the most preferred site), and environmental impacts (measured by an index based on ten components). Thus, the DEA model serves the construction of an index and does not represent any attempt to estimate a production function as such (cf. the statement in Charnes et al. (1990), p. 74, concerning Thompson et al. paper: "Desired here was *not* an empirical production function but the choice of one of the DMUs [sites]." The problem was that running this DEA model five of the six sites turned out as efficient.

Now, the obvious solution to the choice problem would be to calculate the costs in monetary terms and perform a standard economic cost-benefit exercise. However, the claim in Thompson et al. (1986) was that this turned out to be too difficult to carry out, so instead 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.

It is interesting to note that in the calibration of the bounds, detailed data on the build-up of construction costs were used, and also additional information on environmental amelioration costs. Based on the data collected it seems that only a modest effort would be needed to pick the winner, since for the site with both the lowest construction costs and time loss it was stated that the amelioration cost (typically a maximal estimate of environmental costs) “will not constitute a major factor compared to the cost of tunnel construction” (p. 46). Assuming the time loss will be priced identically for all the sites (e.g. based on average wage costs of research) and assuming a modest enough environmental amelioration cost, a direct inspection of the data immediately indicates the same winner (both construction costs and time loss are the smallest, using Table 5 on p. 44 of Thompson et al. 1986) as worked out in the end with some rounds of manipulating the assurance region bounds.

However, a valuable theoretical contribution of Thompson et al. (1986) was to introduce the issue of weight restrictions into DEA and to propose assurance region constraints for weights. The assurance-region approach was further elaborated in Thompson et al. (1990) in a setting with a more normal number of units and within a regular agricultural production function. However, the purpose of weight restrictions was the same as above; to reduce the number of efficient units.

In Dyson and Thanassoulis (1988) there is no mentioning of a production function at all, and neither any reference to Farrell (1957) (in fact there are only four references in the paper). Their concern is a different one from that in Thompson et al. (1986) (not referred to). 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 clearly reveals that they want to attach values beyond a production-function frame of reference to output and inputs.

However, a production function is a technological construct and then neither the issue of importance of a variable (other than if an input is essential in a technical sense for positive outputs) nor issue of value does arise. In production theory (Frisch 1965) one has to keep apart technological considerations concerning transformation of inputs into outputs, and economic considerations of optimising some objective function (e.g., a profit or cost function) subject to the technology constraint (and any other constraints being relevant for the optimisation problem at hand), given input and output prices (in a competitive environment). This separation of technology also holds in the case of producing public services not sold on markets.[7]

Regarding zero weights this reflects the nature of the data at hand, as mentioned in Section 1. A proportional scaling factor for all inputs or outputs, is however, found. But the zero shadow prices, or positive slacks on input and output variables tell us to be careful when interpreting an efficiency score in such circumstances. The general bias is that the efficiency score may have a too high value. However, starting to manipulating weights may not be the way to proceed.

It is also a question of the technological realism of weight restrictions. Dyson and Thanassoulis (1988) introduce a minimum level of the resource per unit of output based on running regressions of the resource on the outputs. However, to state a minimum unit requirement of the input for all outputs is a technological statement that cannot be revealed by such a regression. It also a question if this is compatible with specifying constant returns to scale as done.

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",

[7] However, as observed in Charnes et al. (1978, p. 430, footnote 4) in a very interesting remark, engineering characterisations of technology can be "difficult in public sector programs such as education, public safety, etc., where the meaning of 'technology' is likely to be more ambiguous than in the case of manufacturing in the private sector, and even many service operations."

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. As stated in Section 1 the occurrence of zero weights just reflect the information that can be extracted from the data at hand, given the assumptions imposed on the production possibility set by the analyst.

Bessent et al. (1988) introduce a novel perspective related to the weight-restriction problem, but operate in the input and output space. They start out with referring to Farrell (1957) “in which an empirically derived frontier of relative efficiency, rather than a theoretical production function, is used as the basis for measuring the relative efficiency of units” (p. 785). Referring to the applied literature, the problem they identify is that DEA solutions in some cases “produce efficiency ratings and marginal rates of substitution and productivity that are difficult to interpret and often unacceptable to unit managers” (p. 785). They state that these results have to do with the mix of inputs and/or outputs of inefficient units that are different from any frontier point. Such units are termed “not naturally enveloped inefficient units.” Empirical results referred to indicate that such units constitute the lion’s share of inefficient units (only 2.9 % out of 1132 inefficient units were naturally enveloped, according to Lang et al. 1995, p. 478). The remedy developed is not based on weight restrictions, but on seeking to extend the facets involving not naturally enveloped inefficient units. By extending faces a better envelopment is achieved. The face that is extended is the nearest face to the inefficient unit in question. An elaborate algorithm is set up to find the maximal number of referent units.

However, the approach has its weaknesses. Olesen and Petersen (1996, p. 213) point to the problem created by allowing faces that are of less than full dimension to be candidates for extension, but that it may be impossible to reach a full dimensional facet from such a subsets of efficient units spanning a face of reduced dimensions. Lang et al. (1995, p. 479) point to the fact that it may not be achievable to obtain proper envelopment, and that there may be collinear referents, and envelopment is only measured by the number of linearly independent referents that can be found on an efficient face.

In Charnes et al. (1990), based on an approach presented in Charnes et al. (1989), 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 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 preeminently 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 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.

The approach has interesting aspects form a mathematical point of view, although Olesen and Petersen (1996) criticise the approach because the multipliers for the few chosen efficient units are necessarily not unique, and the choice made as to which estimates of multipliers will be used will have consequences for the transformation of the data for the other units.

However, a more general critique is that 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).

Wong and Beasly (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 they 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.

The two most recent and extensive review articles of weight restriction literature in DEA are Allen et al. (1997) and Pedraja-Chaparro et al. (1997). By accepting the positions in the reviewed literature without questions the papers demonstrate the typical misunderstandings of the philosophical position of weight restrictions and introduction of preferences and values, as commented upon above. In Pedraja-Chaparro et al. (1997, p. 218) the seminal Farrell (1957) Diagram 1 is reproduced. However, the level of understanding of values does not go beyond the level found in the literature that is reviewed.

In the introduction Allen et al. (1997, p. 14) states: "The definition of efficiency in DEA is based on the engineering concept of total factor productivity and is specified as the ratio of the weighted sum of outputs to the weighted sum of inputs of a DMU." To place the concept of total factor productivity within engineering may be a surprise for economists, but more serious in a review paper is the total neglect of the seminal definition of technical efficiency of Farrell (1957).

An interesting new realisation 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 constraint 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. However, neither the implication of this insight for the most appropriate definition of technical efficiency nor the problem of continuing with the Charnes et al. (1978) ratio definition of efficiency in the case of variable returns to scale is commented upon.

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 (2) in Charnes et al. (1978) generally does not lead to the correct evaluation of the relative efficiency of the assessed unit.

Thanassoulis and Allen (1998) point to the rather obvious fact that restricting weights also changes the production possibility set in the input and output space, and that it should be possible to introduce unobserved units that will generate the same change in the estimated production possibility set as introducing constraints in the multiplier spaces.[8] The question is the basis for choice for the unobserved units. A subset of efficient units termed "Anchor points" was defined as the units that should be used as the basis for creating the unobserved unit. The end result is an extension of the efficient frontier that has parallels with the approaches in Bessent et al. (1988) and Lang et al. (1995) (as pointed out in Thanassoulis and Allen 1998).

This idea is followed up in Allen and Thanassoulis (2004). The purpose of introducing unobserved units is sharpened to be to reduce as much as possible units with zero weights (or ε values) in order to make the units "properly enveloped." Anchor points are more formally defined, and an elaborate algorithm in the case of multiple outputs, but a single input and assuming constant returns to scale, is developed.

I will argue that the only 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 shadow prices. But as is evident from Section 3 these properties are facet-specific, so to impose general restrictions seems inappropriate. It would, indeed, be a formidable task to get enough information about properties of each facet. Appealing to market prices, as in Charnes et al. (1990), p. 77, 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 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.

[8] It is interesting to note that already Farrell (1957) introduced artificial units in his estimation problems, having infinity for one input at a time and zero for the other inputs. His purpose was to secure negatively sloped isoquants. Carnes et al. (1978), p. 435, referred to Farrell's "awkward concepts such as 'points at infinity'," but introduce the use of the non-Archimedean ε as a lower constraint on the multipliers. As pointed out in Førsund et al. (2009), and following Thanassoulis and Allen (1998), the two approaches can be made to have an equal impact on the optimal solution.

In three related interesting contributions (Podinovski 2004a; Podinovski 2005; Podinovski 2007) where the trade-off approach is introduced, as mentioned in Section 1, 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 θ in problem (4), 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.

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 defined in (10). (However, trade-offs may be regarded as bounds on such rated.) 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 information about the frontier function rates remains unresolved. Although the university examples used in Podinovski (2004a); (2005); (2007) are quite instructive (and extensive in Podinovski 2007), it is still the question whether the assumption of fixed coefficients called trade-offs of the observed units are technically realistic.

5. Conclusions

It should be recognised that the concept of efficiency in economics is in general defined relative to an efficient production process and does not depend on the method for estimating neither the frontier function nor the efficiency score. However, Farrell (1957) recommended to use a piecewise linear non-parametric best-practice production function when estimating efficiency.

Charnes et al. (1978) started out with an efficiency definition claimed to be based on engineering science formulated as a problem of maximising productivity for each units subject to the constraint that no unit could be more efficient than observed best practice. The weights of the productivities are the endogenous variables. The optimisation problem was transformed to an equivalent linear programming problem. Due to the fundamental duality property of linear programming the apparently different definitions of efficiency are equal.

Empirical applications of DEA showed a prevalence of zero (or ε-level) weights leading to question the consequence for the efficiency estimate based on the ratio definition. The literature on weight restrictions is exclusively based on the efficiency definition in Charnes et al. (1978). It was stated that variables with zero weights had no influence on the efficiency score, in spite of the alleged importance of the variables. This has been one motivation for introducing restrictions on weights. Another empirical result was that often there were too many efficient units. This problem could also be overcome by introducing weight restrictions.

Framing the problem of estimation of efficiency within a linear programming model, this paper has raised serious questions about connecting the constraints on weights to intrinsic economic values of output and input variables. 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. A measure of technical efficiency should not be confused with economic efficiency.

Constraining weights have the implication of extending the production possibility set in input – output space. It may be argued that restricting weights are done in order to incorporate information about the shape of the production function using the connection between ratios of weights and economic concepts of marginal productivity, marginal rate of transformation between a pair of outputs, and marginal rate of substitution between a pair of inputs. However, these properties are specific to each face (and constant on a face) of the estimated frontier production function and it seems an overwhelming task to get such information for all faces. If the trade-off approach of Podinovski (2004a); (2005); (2007) can be realistically based it has some promises. Podinovski pointed out the superiority of the 'technological thinking' over ‘value thinking’. As pointed out in Podinovski and Athanassopoulos (1998) and follow-up papers the equality between the two basic definitions of efficiency may break down when weight restriction are introduced.

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).

Recognising the importance of the shape of the production possibility set there are also efforts on changing the shape of faces by extending them in order to properly envelop inefficient units not being properly enveloped in the initial run and creating many of the problems mentioned above for efficiency measurement. However, extending faces is not easily done in a consistent manner, and there remains a sense of ad hoc about the approach.

The idea of introducing artificial units was already applied in Farrell (1957). He introduced one unit for each input having the value of infinity for each unit in turn, in order to keep his preferred form of isoquants. This idea has resurfaced, and is based on using experts to inspect the resulting frontier in input – output space and work out introducing artificial observations targeted to solve some of the difficulties created by not properly enveloped units (Thanassoulis and Allen 1998; Allen and Thanassoulis 2004) and unduly efficient units (Krivonozhko et al. 2009).

A most pertinent observation is the one made in Olesen and Petersen (1996) that zero multipliers and slacks appear because data do not contain sufficient information to avoid this. Zero shadow prices reflect the data structure relative to the basic axioms of the production possibility set when estimating the efficiency scores.Their remedy for what they call ill-conditioned data sets is to experiment with the degree of aggregation of the data, working on the premise that the higher the degree of aggregation the fewer zero weights will appear. However, this approach does not seem to have been followed up in the literature.

References

Allen R and Thanassoulis E (2004) Improving envelopment in data envelopment analysis. *European Journal of Operations Research* 154, 363-379

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

Bessent A, Bessent W, Elam J, and Clark T (1988) Efficiency frontier determination by constrained facet analysis. *Operations Research* 36(5), 785-796

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

Caves DW, Christensen LR and Diewert WE (1982) The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica* 50, 1393-1414

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

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, Cooper WW, Wei QL and Huang ZM (1989) Cone-ratio data envelopment analysis and multiobjective programming. *International Journal of Systems Science.* 20(7), 1099–1118

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

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 Publ. Comp.

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

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 VE, Utkin OB, Safin MM and Lychev AV (2009) On some generalization of the DEA models. *Journal of the Operational Research Society* 60, 1518–1527

Lang P, Yolalan OR and Kettani O (1995) Controlled envelopment by face extension in DEA. *Journal of the Operational Research Society* 46(4), 473-491

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

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 (1999) Side effects of absolute weight bounds in DEA models. *European Journal of Operational Research* 115, 583-595

Podinovski VV (2001a) Validating absolute weight bounds in data envelopment analysis (DEA) models. *Journal of the Operational Research Society* 52, 221-225

Podinovski VV (2001b) DEA models for the explicit maximisation of relative efficiency. *European Journal of Operational Research* 131, 572-586

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

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

Thanassoulis E and Allen R (1998) Simulating weight restrictions in data envelopment analysis by means of unobserved DMUs. *Management Science* 44(4), 586-594

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

Thompson RG, Langemeier LN, Lee C-H, Lee E and Thrall RM (1990) The role of multiplier bounds in efficiency analysis with application to Kansas farming. *Journal of Econometrics* 46, 93-108

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