Journal of Applied Economics, Vol. VI, No. 1 (May 2003), 157-175

HOW CAN WE USE THE RESULT FROM A DEA ANALYSIS? IDENTIFICATION OF FIRM-RELEVANT REFERENCE UNITS

JONAS MÅNSSON*
Växjö University

Submitted August 2002; accepted February 2003

Two types of guidelines can be obtained from a DEA (data envelopment analysis) analysis. Firstly, the firm can reduce input or increase production according to the DEA results. Secondly, an inefficient firm might be able to identify reference units. This makes it possible for the inefficient firm to, on site, study production that is more efficient, and thereby get information on e.g. efficient organisational solutions. In this study, we focus on how to detect these firm-relevant reference units. While applying the existing methods for identification of reference units, i.e. the *intensity variable* method and the *dominance method*, on a data set concerning booking centres in the Swedish taxi market, shortcomings in these methods were identified. This motivates the development of a new method. This new method, the *sphere measure*, enables an inefficient unit to identify existing and efficient units that have the largest similarity with itself. The identified units will thus be firm-relevant reference units.

JEL classification codes: D24, L25
Key words: reference units, firm-relevant, DEA

I. Introduction

There are two kinds of guidelines that can be provided to firms as a result of a DEA-analysis on technical efficiency.[1] First, one guideline would be

* I would like to thank Prof. Rolf Färe, Oregon State University, Prof. Lars Gunnar Svensson, Lund University and Dr. Ann Veiderpass, Gothenburg University for valuable comments on previous versions of this paper. Correspondence should be addressed to: Department of Economics and Statistics, Växjö School of Management & Economics, Växjö University, S-351 95 Växjö, Sweden.

[1] In data envelopment analysis, DEA (see e.g. Charnes, Cooper and Rhodes, 1978), the

how much a specific unit will be able to reduce its input while still being able to produce the same amount of output. This type of guideline does not take technical or organisational obstacles into consideration.[2] Therefore, a second type of guideline is to identify units that can serve as a reference for an inefficient unit.[3] Relevant reference units make it possible for inefficient units to, on site, study production that is more efficient than its own. This makes it possible to adopt more efficient ways to organise production.

In the literature, two methods are discussed as a means to identify reference units based on the result of a DEA analysis. These are the *intensity variable method* (See Kittelsen and Førsund, 1992) and the *dominance method* (See Tulkens, 1993). We have explored these two methods on a data set concerning the production of booking centre services in Sweden, and identified shortcomings in these methods. In some cases, units, which were defined as reference units for a specific inefficient unit, had little similarity with regard to amount of input used and output produced. Results of this type that are reported to managers will undermine confidence in the DEA method. Furthermore, while investigating the dominance method another shortcoming was identified. For some units, it was not possible to identify a reference unit that dominated the inefficient unit. The identified shortcomings in the existing methods of detecting reference units, for an inefficient unit, motivate the search for a new method. The starting point for this search is to list properties that are desired for reference units. Then we use these properties to construct a measure/method that fulfils these properties.

---

reference technology, is specified as an activity analysis model (see e.g. von Neumann, 1938). The model is also referred to as the non-parametric method (see e.g. Färe, Grosskopf and Lovell, 1985). The input based framework used in this study originates from Farrell (1957) and was later generalised to also cover non-homogeneous production technologies, i.e., allowing for variable returns to scale, by Førsund and Hjalmarsson (1974,1979). The idea was presented in 1974 and implemented in 1979. In Färe, Grosskopf and Lovell (1983), the framework was further generalised to cover multiple output and different disposability assumptions.

[2] For example, a small unit may find it efficient to handle administrative issues manually, while large units computerise.

[3] This is unlikely to happen in a competitive environment, but in e.g. the public sector, providing this information to others may not be a problem.

The outline of this study is as follows. In Section II, we will state the framework used in the study. We start by set up the DEA problem and presenting a list of desired properties. These properties are as follows. A reference unit should always exist, the reference unit should be efficient, the reference unit should be an existing unit (i.e. excluding hypothetical reference units such as convex combinations of existing units), and finally a reference unit should be as similar as possible to the inefficient unit. Data is presented in Section III. In Section IV, we first evaluate the existing methods with respect to the desired properties presented in Section II. As mentioned above, we could show that in some cases, designated reference units had little similarity with the inefficient unit. In the case of the dominance method, we could also show that reference units in some cases did not exist. We therefore introduce a new method, the s*phere measure*, which is constructed so that it will fulfil the desired properties. The method will guarantee the existence of a unit, chosen among existing efficient units so that it will minimise the Euclidean norm between the reference unit and the inefficient unit, i.e. has the largest similarity. In Section V, the results are summarised and some concluding remarks are stated.

II. Framework

A. Measuring Efficiency with DEA

Since the aim of this study is to state desired properties of a reference unit, as a result of a DEA analysis, we first need to set up the DEA problem. Let there be $k = 1,\ldots, K$ observations, x_{kn} inputs $n = 1,\ldots, N$, and y_{km} outputs, $m = 1,\ldots,M$. The vector to be enveloped for observation k is then $(x_k, y_k) = (x_{k1},\ldots, x_{kN}, y_{k1},\ldots, y_{kM})$. Then the programming problem to be solved for a unit k' is as follows

$$TE(x_{k'}, y_{k'}) = Min\ \lambda_{k'} \tag{1}$$

s.t.

$$i) \qquad \sum_{k=1}^{K} z_k y_{km} \geqq y_{k'm},\ m = 1,\ldots, M$$

$$ii) \quad \sum_{k=1}^{K} z_k x_{kn} \leqq \lambda_{k'} x_{k'n},\ n = 1,...,N$$

$$iii) \quad \sum_{k=1}^{K} z_k = 1$$

$$iv) \quad z_k \geqq 0,\ k = 1,...,K$$

where $\lambda_{k'}$ is the efficiency score to be calculated. Since an input based framework is used, the minimum of $\lambda_{k'}$ equals the largest possible contraction of the input vector, such that the unit still remains in the reference technology. We also assume strong disposability of both inputs and outputs and a variable return to scale technology. The latter is given by restriction *iii*.

B. Desired Properties of Reference Units

Before stating and discussing desired properties of a reference unit, some definitions and notations have to be made. First, denote the set of all observed units by $\mathbf{K} = \{1,\dots, k, K\}$. The set of reference units for a specific unit k is denoted $\Re e_k$, i.e. if unit j is a reference unit for unit k, then $j \in \Re e_k$. Finally, given an input requirement set $L(y)$, we can define the isoquant of this input requirement set as $Isoq\ L(y) = \{x : x \in L(y), \lambda x \notin L(y)$ for all $\lambda \in [0,1]\}$. Given the definitions and notations above, we will state desired properties and subsequently discuss them.

Table 1. Desired Properties of a Reference Unit/s

	Property
1	$\Re e_k \neq \varnothing$
2	If unit $j \in \Re e_k$ then $x_j \in Isoq\ L(y)$
3	If unit $j \in \Re e_k$ then unit $j \in \mathbf{K}$
4	If unit $j \in \Re e_k$ then there cannot exist another unit i, $x_i \in Isoq\ L(y)$, such that $\lVert ik \rVert < \lVert jk \rVert$

A first property is that at least one possible reference unit should exist, i.e. $\Re e_k \neq \emptyset$. This property might seem redundant, but as will be discussed later, one of the existing methods may produce results where a reference unit does not exist.

Since a goal for all economic activity is the efficient use of resources, the second property we claim for a reference unit is that it should be efficient. This is given by the second property that states: if unit j is a reference for an inefficient unit k, i.e., $j \in \Re e_k$ then it is impossible to contract the input vector of unit j, while still being able to produce the same amount of outputs, i.e. $x_j \in Isoq\, L(y)$.

Further, the aim of using a reference unit is that it should be possible for an inefficient unit to study the production of the reference unit on site. The third property states that if unit j is to be a reference unit for an inefficient unit k, i.e. $j \in \Re e_k$, unit j has to be observable, i.e. $j \in \mathbf{K}$. Thus, property 2 excludes convex combinations of existing units.

So far we have excluded all other inefficient units and convex combinations of existing efficient units from the possible reference set. However, we are still left with a considerable amount of possible units. From a practical point of view, to make an impact on firms trying to become more efficient, we need to guide them to reference units that in some sense are similar to their own firm. The term similarity is not easy to define since two units can be similar/dissimilar in many different dimensions.[4] However, since DEA analysis is an analysis of production and researchers are likely to at least have information about production data, we therefore define similarity as producing a similar amount of outputs and use a similar amount of inputs. To define similarity in a multidimensional framework, we need a measure that is able to take multidimensionality into consideration. The Euclidean norm is one such measure and will here be used as a measure of similarity. Further, we will claim that the most similar unit among possible reference units is most suitable reference unit. Thus, the forth desired property of a reference unit j is that another possible reference unit $i \in \Re e_k$ there should not exist, such that the distance between unit i and the inefficient unit k is smaller than the distance between unit j and unit k, i.e. $\|jk\| < \|ik\|$ for all i.

[4] E.g., two units can be similar with respect to location, education of management, gender representation etc.

Given the properties above, we now turn to empirically explore these properties related to a data set. We start by exploring the two existing methods, the *intensity variable method* and the *dominance method*, and finally we introduce a new method labelled the *sphere measure*.

III. Data

The data in this study concerns production of booking centre services in the Swedish taxi market. The data was collected and confirmed on site at the booking centres during a three-week period in March 1994 and later used in Althin, Färe and Månsson (1994).[5] The production of booking centre services consists of two outputs. The first output is a measure of directly mediated service (*Y1*), i.e. a person orders a taxi and the booking centre immediately mediates the order to a taxi vehicle. The production of the second output, number of co-ordinated and mediated services (*Y2*), is carried out in two steps. The first step is that a person orders a taxi. The order will be co-ordinated with other orders, either by placing more than one customer in the taxi vehicle or by re-directing the taxi vehicle to minimise non yielding transportation. After co-ordination, the order is mediated to the taxi vehicle.

The inputs are:

X1: Number of hours worked annually by personnel directly involved with booking and mediation.
X2: Numbers of hours worked annually by administrative staff.
X3: Number of telephone lines to the booking centre. This will serve as a measure of technical capacity.
X4: Square meters of floor space used for booking services.
X5: Square meters of floor space used for administration.
X6: Value of purchased services in Swedish kronor (SEK).

Descriptive statistics on input and output are presented in Table 2.

A few comments have to be made concerning the data. One can see that there are booking centres that only produce one of the outputs. This can be explained by the fact that the data covers both privately owned and publicly owned

[5] For a more extensive discussion on booking centre production, see Månsson (1996).

Table 2. Descriptive Statistics on Inputs and Outputs for the Production of Booking Centre Services (N = 30)

	Mean	Std. Dev.	Min	Max
Output				
Directly mediated services (Y1)	176,720	208,677	0	1,000,000
Co-ordinated and mediated services (Y2)	77,701	99,041	0	400,000
Input				
Hours worked with booking - mediation (X1)	11,226	10,302	979	54,136
Hours worked with administration (X2)	3,648	4,410	0	20,976
Telephone lines to the booking centre (X3)	10	7.12	1	28
Floor space used for the booking services (X4)	35	35.7	6	200
Floor space used for administration (X5)	40	56.7	0	300
Value of purchased services in SEK (X6)	99,000	271,753	0	1500,000

booking centres. One of the objectives with introducing publicly owned booking centres was to increase the number of co-ordinated services. This explains why *Y1* for some booking centres is zero. On the other hand, the most likely way to administrate an order during the period when the Swedish taxi market was regulated was to mediate the order at the same moment a customer placed the order in the booking centre. Some privately owned booking centres still apply this system, and thereby do not allocate resources to co-ordinate services. This explains the zero value for *Y2.* Zero input values can partly be explained by the fact that some booking centres do not have any administrative staff, instead they buy administrative services. This is most likely to happen in the case of small booking centres.

IV. Empirical Investigation

We first present here the computed efficiency scores. Thereby all units that fulfil property 2 and property 3, i.e., all existing efficient units, are identified. We then apply the existing methods, *dominance* and *intensity variable method*, on the data presented in Section III. As will be seen, both existing methods have some shortcomings as regards desired properties. We therefore propose a new method, which will be labelled the *sphere measure*.

A. Identification of Existing and Efficient Units

The framework presented in Section II was used to compute the efficiency scores. The results of these computations are presented in Table 3.

Table 3. Technical Efficiency, Variable Returns to Scale

Unit no.	Efficiency score	Unit no.	Efficiency score	Unit no.	Efficiency score
1	1.000	11	1.000	21	1.000
2	0.875	12	0.980	22	0.663
3	0.722	13	0.523	23	0.490
4	1.000	14	0.748	24	1.000
5	1.000	15	0.769	25	0.797
6	1.000	16	1.000	26	1.000
7	1.000	17	1.000	27	0.793
8	0.584	18	0.901	28	0.694
9	0.806	19	0.950	29	1.000
10	0.641	20	1.000	30	0.758

As seen in the Table, thirteen units are efficient. The minimum efficiency is 0.49 for unit number 23. This means that unit 23 would have to decrease its inputs by 51 percent in order to become efficient. The mean efficiency score is 0.86, i.e. 14 percent inefficiency, and the standard deviation is 0.16. All units that are technically efficient, i.e. have an efficiency score equal to one, fulfil property 2 and are thus potential reference units. Further, they also fulfil property 3, i.e. are existing units.

B. Existing Methods for Detecting Reference Units

B.1. Intensity Variables

When the non-parametric method is used to compute technical efficiency, inefficient units are compared to a convex combination of efficient units. By investigating the value of the individual intensity variables (z_k), obtained when solving the efficiency problem presented in equation (1), it is possible to identify those units that are used in the construction of the efficiency frontier. According to Kittelsen and Førsund (1992), p.302, this information can be used to select a reference unit among the efficient units.[6]

In Table 4 below, the values of the non-zero intensity variables are presented for the inefficient units. These results can be used to provide the inefficient unit information on which efficient unit it is compared to. For example, the inefficient unit 9 is compared to efficient units 1, 7, 11 and 29. According to the values of the intensity variable, efficient unit 11 is the most relevant reference unit, since it has the highest value on the intensity variable (0.754).

One problem with this method occurs when the most influential unit has very little similarity with the inefficient unit.[7] One way of handling this drawback would be to report all units with non-zero intensity variables. It does not solve the problem, but it will provide the inefficient units with alternative units to be compared with. Another way is to determine some criteria for similarity and investigate if the designated unit is the most similar reference unit.

[6] When using the approach suggested by Kittelsen and Førsund, it is possible that more than one reference unit exists. This will be the case if two, or more units have the same value on their intensity variables.

[7] As can be seen in the Appendix, unit 11 is using much less input and produces much less output in each dimension. My experience is that reporting this type of information back to managers will induce suspicion and undermine creditability of the method, since managers will not see unit 11 as a relevant reference unit.

Table 4. Inefficient Units, Units Used in the Reference Frontier for the Inefficient Unit (Frontier Unit), and the Values of the Intensity Variable

Inefficient unit no.	Frontier unit. no	Intensity variable	Inefficient unit no.	Frontier unit no.	Intensity variable	Inefficient unit no.	Frontier unit no.	Intensity variable
2	7	0.2645		20	0.2246		29	0.0588
	11	0.0990		21	0.1182	23	5	0.0519
	24	0.5247		29	0.4906		11	0.3453
	29	0.1118	13	4	0.5563		16	0.4375
3	16	0.6941		11	0.2209		24	0.0255
	24	0.0144		16	0.1943		29	0.1399
	29	0.2915		29	0.0284	25	7	0.0261
8	7	0.1815	14	4	0.1060		11	0.3035
	21	0.5108		16	0.8940		16	0.4711
	29	0.3078	15	7	0.2194		21	0.1697
9	1	0.0779		16	0.6719		29	0.0297
	7	0.0069		29	0.1087	27	5	0.8044
	11	0.7540	18	16	0.5201		24	0.1956
	29	0.1612		24	0.3517	28	6	0.2496
10	1	0.1005		29	0.1282		7	0.5638
	7	0.0174	19	4	0.0146		29	0.1867
	11	0.2852		16	0.8025	30	11	0.7108
	21	0.4904		29	0.1829		16	0.2155
	29	0.1064	22	4	0.3339		24	0.0193
12	11	0.1666		16	0.6073		29	0.0543

Note: As can be noted, the efficient units 17 and 26 are not used as reference for any inefficient unit. The most likely explanation for this is that both these units are unique, in the sense that they are only compared with each other. They are located on either the vertical or the horizontal line segment in Figure 1.

In this study, we use the difference in the Euclidean norm to *indicate* similarity. The Euclidean norm measures the distance between units. The norm between unit i and k is here defined as:

$$\| ik \| = \sqrt{\sum_{n=1}^{N} \left(\frac{x_{ni}}{\overline{x}_n} - \frac{x_{nk}}{\overline{x}_n}\right)^2 + \sum_{m=1}^{M} \left(\frac{y_{mi}}{\overline{y}_m} - \frac{y_{mk}}{\overline{y}_m}\right)^2} \quad (2)$$

where $\overline{x}_n$ and $\overline{y}_m$ is the mean of n / m.[8]

The criteria we use is that if the norm between unit j and unit k is smaller than the norm between another unit i and unit k, i.e. $\|ik\| > \|jk\|$, then unit j is more similar to k than unit i is to unit k, and thereby also a more relevant reference unit. We have computed the Euclidean distance between unit 11 and all other observed efficient units and that result is presented in Table 5.

Table 5. Euclidean Distance between Unit No. 11 and all Other Observed and Efficient Units

Efficient unit	Unit. No. 9	Efficient unit	Unit. No. 9
1	2.32	17	6.67
4	2.45	20	2.60
5	2.30	21	2.69
6	2.00	24	2.75
7	1.98	26	2.44
11	2.60	29	6.62
16	2.47		

As shown in the Table there is a unit that have larger similarity to unit 9 than the by intensity variable method detected unit 11.[9] We can thus conclude that that the intensity variable does not fulfil the desired property 4.

[8] The data is normalised since the norm otherwise will be dependent on how the data is measured.

[9] The difference in each input and output dimension, between unit No. 9 and unit No. 7 is reported in the Appendix.

B.2. Dominance

There is one major critique of the non-parametric, or DEA framework presented above. When computing the efficiency score, the inefficient units are compared with convex combinations of efficient units, instead of existing units. As a consequence of this, Tulkens (1993) presented the idea of dominance, which in turn has its roots in Pareto efficiency.[10] In a multiple input, multiple output framework dominance can be defined either from the input, or the output side. Following Tulkens (1993), input dominance is defined as:

Definition: A unit *k* input dominates k', if and only if

$$y_{km} \geqq y_{k'm} \;,\; m = 1,...,M \;\; and \;\; x_{kn} \leq x_{k'n} \;,\; n = 1,...,N$$

That is, unit *k* input dominates *k'*, if unit *k* produces more or equal amount of output compared to *k'* $(\geqq)$ and uses less input in *at least* one dimension $(\leq)$. An alternative version of dominance is strict dominance, taking both inputs and outputs into consideration at the same time.

Definition: A unit *k* strictly dominates unit *k'*, if and only if

$$y_{km} > y_{k'm} \;,\; m = 1,...,M \;\; and \;\; x_{kn} < x_{k'n} \;,\; n = 1,...,N$$

That is, unit *k* strictly dominates unit *k'*, if unit *k* produces more output and uses less input in all dimensions. This means that if unit *k* strictly dominates unit *k'*, then unit *k* also input and output dominates unit *k'*.

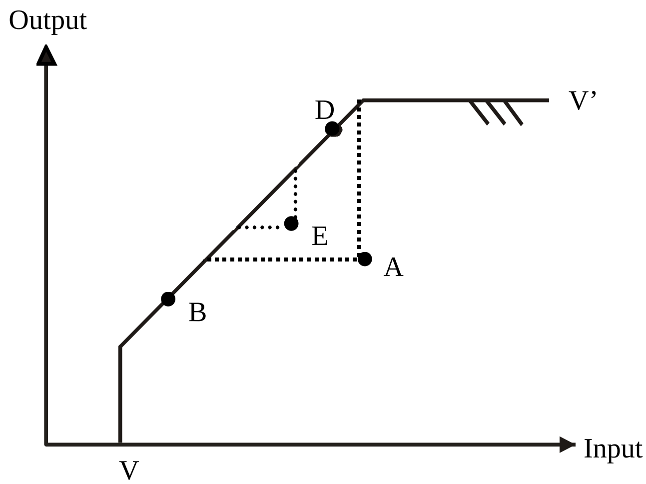
As noted by Tulkens (1993), p.191, identification of a dominant unit gives the efficiency score credibility, since it identifies an observed reference unit, instead of a convex combination of existing units.[11] Dominance and a problem with the method are illustrated in Figure 1.

---

[10] In Tulkens (1993), the author uses the idea of dominance to construct a new reference technology, labelled Free Disposal Hull reference technology (FDH). It should be noted that in this study, we apply the ideas of dominance, given the convexity assumptions of the reference technology, i.e. we do not use the FDH reference technology.

[11] Output dominance is defined analogously, with strong inequality in at least one output dimension.

Figure 1. Illustration of Dominance



In this Figure, unit *A* and unit *E* are inefficient. It is clear that unit *A* is strictly dominated by the efficient unit *D*, since unit *D* uses less input and produces more output than unit *A*. A problem arises, if a situation illustrated by the inefficient unit *E* occurs. Even though unit *E* is inefficient, it is neither dominated by unit *B* nor *D*.[12] Unit *E* produces less output, but at the same time uses less input, compared to unit *D*. The opposite is true when comparing with unit B. Thus, this method may result in a situation where the dominant subset is empty. Dominating references were found for two units for the data used in this study. The efficient unit No. 7 dominated both the inefficient units No. 3 and No. 10. For all other inefficient units, the dominant sub-set was empty, i.e. $\Re e_k \neq \emptyset$. This result was not unexpected, since the model on which the computations were based has as many as 8 dimensions: 2 output dimensions and 6 input dimensions. The more dimensions used in the model, the less likely it is that the dominant subset is non-empty. Thus, the *dominance method* might not fulfil property 1 or property 4.

---

[12] If the FDH reference technology was used, point E had been considered efficient.

C. The Sphere Measure

In Section IV.B we have demonstrated the intensity variable method and the dominance method and we have identified shortcomings in both methods. We therefore propose a new method with the objective of identifying firm-relevant reference units that fulfil the desired properties listed in Section II.

The idea of the *sphere measure* is rather straightforward. For an inefficient unit, a sphere with radius r is defined. The radius of the sphere is then extended until the sphere covers the inefficient unit and at least one efficient unit. The unit that first appears in the interior of the sphere is considered to be the reference unit for the inefficient unit.[13] Moreover, the length of the radius is a measure of how close the inefficient unit is located to the reference unit.

First, denote the subset of efficient observations $\mathbf{S} \subseteq \mathbf{K}$. The subset $\mathbf{S}$ contains all efficient units from the set of all units, $\mathbf{K}$. For an inefficient unit k, and an efficient unit $s \in \mathbf{S}$, the radius of the sphere is defined and computed as:

$$r_{ks} = \sqrt{\sum_{n=1}^{N}\left(\frac{x_{ns}}{\bar{x}_n} - \frac{x_{nk}}{\bar{x}_n}\right)^2 + \sum_{m=1}^{M}\left(\frac{y_{ms}}{\bar{y}_m} - \frac{y_{mk}}{\bar{y}_m}\right)^2} \qquad (3)$$

where r_{ks} is the radius of the sphere. $\bar{x}_n$ and $\bar{y}_m$ denotes the mean of inputs and outputs.

If we let the radius of the sphere increase until it contains the inefficient observation k and the efficient observation s, we can define the reference unit for the inefficient unit k as:

Definition: The efficient observation s is a reference to the inefficient observation k if

$$r_{ks'} = \text{minimum } r_{ks}, \ \forall s \in \mathbf{S}$$

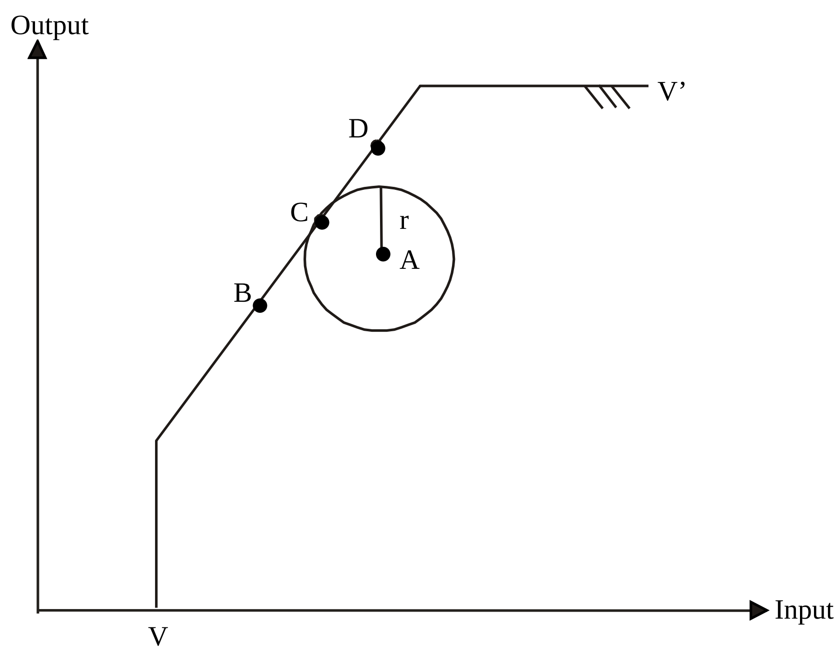
min r_{ks} is thus the smallest distance between all efficient units and the evaluated

---

[13] It is possible that more than one reference unit exists. This will be the case if two, or more units have the value of the *sphere measure.*

inefficient unit *k*. The expression is interpreted as the minimum radius of the sphere, such that the sphere contains at least one efficient unit and the unit *k*. The solution to the minimising problem identifies the efficient unit that is located closest to the inefficient unit, measured by the Euclidean distance. The *sphere measure* is illustrated in Figure 2.

Figure 2. Illustration of the Sphere Measure



In this Figure, units *A*, *B*, *C* and *D* are the observed units, thus **K** = {*A*, *B*, *C*, *D*}. Among these units, *A* is inefficient, while *B*, *C*, *D* are efficient, thus **S** = {*B*, *C*, *D*}. When the radius, *r*, increases, unit *C* will be the first unit to appear within the sphere. The efficient unit *C* is then defined as a reference to unit *A*.[14] The result of the computation of the *sphere measure* for the data is presented in Table 6.

[14] Note that since the *sphere measure* searches for the most similar unit in all directions, it is possible that the selected reference unit use more input in one or more than one dimension. Depending on input prices this *could*, as in the intensity variable method, result in a situation of increased cost. To exclude this situation, information about input prices is necessary.

Table 6. Descriptive Statistics of the Sphere Measure

Inefficient unit no.	Mean r_{ks}	Std Dev. r_{ks}	Min. r_{ks}	Reference unit	Inefficient unit no.	Mean r_{ks}	Std Dev. r_{ks}	Min. r_{ks}	Reference unit
2	4.23	1.01	1.90	24	18	3.17	1.47	1.96	4
3	3.27	1.61	1.88	6	19	2.83	2.21	1.50	16
8	4.19	1.44	2.06	4	22	3.40	1.69	1.61	4
9	3.11	1.60	1.98	7	23	2.82	2.00	1.57	5
10	3.10	2.22	1.61	1	25	2.46	2.76	0.48	16
12	9.51	0.71	7.80	24	27	4.38	1.39	3.32	24
13	4.49	1.27	2.32	4	28	15.20	0.68	13.69	26
14	2.68	2.57	1.03	1	30	2.62	2.58	0.79	1
15	2.76	2.00	1.28	6					

Note: The Min. Radius represents the distance between the inefficient unit and the closest located efficient unit.

For the data used in this study, it was also possible to identify a unique reference unit with the *sphere measure*. Another appealing feature with the *sphere measure* is that a measure of proximity is also obtained. This makes it possible to evaluate the relevance of the identified reference unit. As can be seen from Table 6, the *sphere measure* varies from 0.48 to 13.69. This also indicates that some detected reference units are better suited than others.

V. Conclusions

The objective of this study has been to provide guidelines on what properties one can expect from a reference unit and also how these reference units could be detected. There is no doubt that reference units can play an important part when the results from an efficiency study are implemented in the investigated industry. Relevant reference units make it possible for an inefficient unit to study, on site, production that is more efficient than its own. This makes it possible for an inefficient unit to adopt a more efficient way to organise its production. The main question for this study has been how we can identify relevant reference units for a firm.

The literature suggested two methods, the *intensity variable method* and *the dominance method*. These methods were used on a data set on booking centre services in Sweden and some shortcomings were identified. Firstly, some pointed out reference units had little similarity with the inefficient unit. Secondly, when using the *dominance method*, no reference unit existed. These shortcomings motivate the search for a new method. To derive the new method, we started with a list of properties that are desired for a reference unit. A reference unit should always exist, the reference unit should be efficient, the reference unit should be an existing unit and finally, the reference unit should be similar to the inefficient unit. Given this list of properties; a new method labelled the *sphere measure* was developed. The idea with the *sphere measure* is to define a sphere around an inefficient unit and then expand the radius of the sphere until it contains the inefficient unit and at least one efficient unit. The unit that first appears in the sphere is then chosen as a reference unit. One advantage with the *sphere measure* is that it is constructed to fulfil all desired properties. In Table 7, the result concerning fulfilment of the four properties, with respect to methods are summarised.

By using the *sphere measure*, the efficient unit that has the largest similarity, measured by the Euclidean distance, is identified as a reference.

Table 7. Comparing Different Methods to Detect Reference Units

	Property	Dominance	Intensity	Sphere
1	$\Re e_k \neq \varnothing$	No	Yes	Yes
2	If unit $j \in \Re e_k$ then $x_j \in Isoq\,(Ly)$	Yes	Yes	Yes
3	If unit $j \in \Re e_k$ then unit $j \in \mathbf{K}$	Yes	Yes	Yes
4	If unit $j \in \Re e_k$ then there cannot exist another unit i, $x_i \in Isoq\,(Ly)$, such that $\|ik\| < \|jk\|$	No	No	Yes

Appendix

Comparing the Input and the Output Vectors between Unit 7, Unit 11 and Unit 9

	Unit No. 9	Unit No. 11	Difference 9 vs.11	Unit No. 7	Difference 9 vs.7
Output					
Directly mediated services (Y1)	170,000	8,140	-161,860	150,000	-20,000
Co-ordinated and mediated services (Y2)	75,000	12,210	-62,790	150,000	75,000
Input					
Hours worked with booking - mediation (X1)	18,651	2,268	-16,383	5,017	-13,634
Hours worked with administration (X2)	1,049	0	-1,049	105	-944
Telephone lines to the booking centre (X3)	11	1	-10	5	-6
Floor space used by the booking services (X4)	55	9	-46	27	-28
Floor space used for administration (X5)	30	9	-21	10	-20
Value of purchased services in SEK (X6)	27,000	16,000	-11,000	70,000	43,000

VI. References

Althin, R., Färe, R. and J. Månsson (1994), *Beställningscentralers Kostnadseffektivitet: Effektivitet vid Produktion av Beställningscentraltjänster med Hänsyn tagen till Stödjande Organisation, (Efficiency in Production of Booking Centre Services, Including the Overhead Organisation)*, Kommunikationsforskningsberedningen, Stockholm, Sweden (In Swedish).

Charnes, A., Cooper, W.W and E. Rhodes (1978), "Measuring the Efficiency of Decision Making Units," *European Journal of Operations Research* **2**: 429-444.

Färe, R., Grosskopf, S. and C.A.K. Lovell (1983), "The Structure of Technical Efficiency," *Scandinavian Journal of Economics* **85**: 181-190.

Førsund, R.F. and L. Hjalmarsson (1974), "On the Measurement of Productive Efficiency," *Scandinavian Journal of Economics* **76**: 141-154.

Førsund, R.F. and L. Hjalmarsson (1979), "Generalized Farrell Measure of Efficiency: An Application to Milk Processing in Swedish Dairy Plants," *Economic Journal* **89**: 294-315.

Farrell, M.J. (1957), "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, *Series A. General* **120** (Part 3): 253-281.

Kittelsen, S.A.C. and R.F. Førsund (1992), "Efficiency Analysis of Norwegian District Courts," *Journal of Productivity Analysis* **3**: 277-306.

Månsson, J. (1996), "Technical Efficiency and Ownership: The Case of Booking Centres on the Swedish Taxi Market," *Journal of Transport Economics & Policy* **30**: 83-93.

Tulkens, H. (1993), "On FDH Efficiency Analysis: Some Methodological Issues and Application to Retail Banking, Courts and Urban Transit," *Journal of Productivity Analysis* **4**: 183-210.

von Neumann, J. (1938), "Über ein Ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes," in K. Menger, ed., *Ergebuisse eines Mathematischen Seminars*, Vienna. Translated by Morgenstern (1945-46), *Review of Economics Studies* **13**: 1-9.