International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Dealing with Interval Scale Data in Data Envelopment Analysis

Halme, M., Joro, T. and Koivu, M.

IIASA Interim Report
December 1998

IIASA
International Institute for Applied Systems Analysis • A-2361 Laxenburg • Austria
Tel: +43 2236 807 • Fax: +43 2236 71313 • E-mail: info@iiasa.ac.at • Web: www.iiasa.ac.at

INTERIM REPORT IR-98-112 /December

Dealing with Interval Scale Data in Data Envelopment Analysis

Merja Halme (mhalme@hkkk.fi)
Tarja Joro (joro@hkkk.fi)
Matti Koivu (koivu@hkkk.fi)

Approved by
Pekka Korhonen (korhonen@iiasa.ac.at)
Leader, *Decision Analysis and Support Project*

Contents

Abstract

This paper considers the problem of interval scale data in the most widely used models of Data Envelopment Analysis (DEA), the CCR and BCC models. Radial models require inputs and outputs measured on the ratio scale. Our focus is on how to deal with interval scale variables especially when the interval scale variable is a difference of two ratio scale variables like profit or the decrease/increase in bank accounts. Using these ratio scale variables as variables in a DEA model we suggest radial models. An approach to how to deal with interval scale variables when we relax the radiality assumption is also discussed

Keywords: Efficiency Analysis, Data Envelopment Analysis, Interval Scale Variables, Negative Variables

Acknowledgments

The research was supported, in part, by grants from the Foundation of the Helsinki School of Economics and Business Administration and Academy of Finland. The authors wish to thank Professor Pekka Korhonen, IIASA, for valuable comments.

About the Authors

Merja Halme is a senior researcher and Tarja Joro and Matti Koivu doctoral students at the Helsinki School of Economics and Business Administration. Tarja Joro was a participant of the Young Scientist Summer Program in 1998.

Dealing with Interval Scale Data in Data Envelopment Analysis

Merja Halme
Tarja Joro
Matti Koivu

1. Introduction

After its introduction (Charnes, Cooper and Rhodes, 1978) Data Envelopment Analysis (DEA) has gained wide popularity producing numerous applications reports as well as articles in scientific journals. In many applications interval scale variables like profit and changes in different variables (like sales, loans etc.) have been used as inputs and/or outputs.

However, the obviously most widely used DEA models (the CCR model with constant returns to scale and BCC model with variable returns to scale) require ratio scale, preferably nonzero data. Data on the interval scale does not allow division, the zero point is not defined and only distances can be calculated. This means that in the CCR and BCC models all the inputs and outputs should be (strictly) positive. The ratio of virtual efficient inputs/outputs and observed inputs/outputs plays a central role in the calculations. In the literature there have been various approaches to deal with negative data. It has, however, remained mostly unobserved that negative data are often observations of variables measured on the interval scale.

It seems that usually – and especially in the applications we have run into (see Section 2.2) - the interval scale variables used in DEA applications are a result of the deduction of two ratio scale variables. Based on this finding we propose - when this is the case - that the original interval scale variable should be replaced by those two ratio scale variables. Moreover we suggest that the weights (prices) of the variables in the resulting DEA optimization model could be set equal. We prove that the approach proposed maintains efficient units always efficient (See Section 3.3).

In this paper a spectrum of approaches to deal with interval scale data is discussed. On one end of it we see the radial model we propose and on the other end the simple diagnosis of units as efficient or inefficient without providing any efficiency score. Somewhere in the middle we see the approaches introduced in Section 4, the general weights procedure as well as procedures not producing a score, but which in addition to the diagnosis as efficient/inefficient produce some more information of the inefficient units. Throughout the paper our main focus is on the CCR and BCC models.

The paper is organised as follows. Section 2 discusses interval scale data and DEA and illustrates the approach proposed in this paper. Section 3 introduces formally our idea. Section 4 discusses other approaches: DEA models with general weights as well as procedures that do not produce scores. Section 5 provides a numerical example and Section 6 presents the conclusions.

2. Translation Invariance and Interval Scale Data

2.1 Translation invariance of a model

Adding a sufficiently large positive constant to the values of the problematic variable has served often as a remedy when there are negative observations in DEA applications. Translation invariance of DEA models has been studied in various papers during the 90´s. A translation invariant model is such that "an affine displacement of data does not alter the efficient frontier" (Ali and Seiford, 1990). The BCC model has been found translation invariant (Ali and Seiford, 1990). However, if the efficiency scores should in addition not be affected, then the BCC output oriented model allows a translation of inputs and the input oriented model of outputs (Lovell and Pastor 1995, Pastor, 1996).

The translation invariance is connected with the convexity constraint of and thus the CCR model does not fulfil this requirement. The aim of this study is to develop a way to deal with interval scale (possibly negative) inputs and outputs without any restrictions for the orientation of the BCC and CCR models. Also we want to retain the original interpretability of the inefficiency scores owing to the radial character of the model. Above all, we want to remain as close as possible to the original spirit of the classical DEA. The proposed approach could be especially welcome for CCR model users and those BCC model problems where both input and output variables include negative observations or where the natural orientation of the model (output/input) does not allow the translation of an interval scale variable (output/input).

2.2 Interval scale data problems

As already mentioned, DEA requires ratio scale data. We did not find so far explicit discussion of the problems caused by data on the interval scale. Negative data values were, however, observed frequently. Our core observation has been that the variables with negative observations we encountered have been - thus far - a result of a deduction of two ratio scale variables. Pastor (1994) lists the following examples of variables in the DEA literature with negative values: increment of time deposits and of demand deposits, rate of growth of gross domestic product per capita, profit and taxes (difference of income items and "cost" items). All these variables mentioned are differences of two ratio scale variables. He also mentions negative taxes, which we feel, however, is not relevant here owing to the timing aspect (the taxes returned date from an earlier period).

The first principle we want to point out even here is the necessity of having the correct variables in the analysis and the necessity of easy interpretation. Also when deciding about the variables one should bear in mind what the radiality of the model used means, that efficiency scores produced should reflect the improvement needs of inputs and/or outputs.

In the sequel we restrict ourselves to the apparently common case where the variable is a difference of two ratio scale values. In general, the problem of how to cope in radial models with variables that result from the division of an interval scale variable (say, the absolute change in the number of bank accounts) by a ratio scale variable (original number of bank accounts) has not been discussed much. Lovell's (1995) solution to the problem is interestingly very much in line with the approach in section 4.1. He has resorted to a transformation in the output variables which relates them to their range in the data. This abolishes the problem of negative values. However, he remarks that the FDH scores calculated for inefficient units are not invariant to the translation.

We suggest that the original interval scale variable should be replaced by the two ratio scale variables. Depending on the character of the original variable, the new variables should be interpreted either as one input and one output like

profit = income-cost,

where both the income and the cost are variables the magnitude of which is under control or one discretionary input (output) and one non-discretionary input (output) like

increment in bank accounts =
number of accounts now - number of accounts at time point t..

The current number of accounts would be a normal output variable and the comparison value would be treated as a non-discretionary input á la Banker and Morey (1986). Moreover we suggest that the weights of these "twin" variables in the DEA dual model should be set equal. This approach was adopted also in a DEA application where the quality of perinatal care was measured by setting the number of babies at risk surviving an output variable and number of babies at risk an input variable (Thanassoulis, Boussofiane and Dyson, 1995). The actual output was the deaths of babies at risk to be minimized. The model thus can be seen as a weight restriction model. According to the classification of Thompson et al (1990) concerning relative weight restriction models the model proposed can be seen as an Assurance Region (AR) model, more closely of type ARII.

Using the proposed procedure we maintain the applicability of the radial model. When the interval scale variable takes negative values in the data it seems quite natural to proceed in the way proposed. However, even in the case when the values of the variable happen to be positive in the data we strongly suggest the approach among other things for the quite obvious reason that division on the interval scale is not allowed.

3. An Approach Decomposing the Interval Scale Variable into Two Ratio Scale Variables

3.1 The case when both the two new variables can be considered as objectives

Assume we have *n* DMUs each consuming *m* inputs and producing *p* outputs. Let $\mathbf{X} \in \Re_+^{m \times n}$ and $\mathbf{Y} \in \Re_+^{p \times n}$ be the matrices, consisting of nonnegative elements, containing the observed input and output measures for the DMUs. We denote by $\boldsymbol{X}_i$ (the *i*th row of $\boldsymbol{X}$) the *i*th input values and by x_{ij} the quantity of input *i* consumed by DMU*j*, assumed to be nonnegative. A similar notation is used for outputs. Furthermore, we denote $\boldsymbol{1} = [1, ..., 1]^T$.

For the sake of symmetry we now introduce the combined DEA problem with variable returns to scale (BCC model) where both outputs are maximized and inputs are minimized (see, e.g., Joro et al. 1998). The CCR model can be obtained by dropping the convexity constraint $\mathbf{1}^T\lambda = 1$. Because we want to end at a linear model we use the directional distance function (see Chambers, Chung and Färe, 1996).

Table 3.1 The combined BCC Model

Combined BCC Primal (BCC_P - C)	Combined BCC Dual (BCC_D – C)
max $\sigma + \varepsilon \mathbf{1}^T(\boldsymbol{s}^+ + \boldsymbol{s}^-)$ (3.1a) s.t $\boldsymbol{Y}\lambda - \sigma \boldsymbol{y}_0 - \boldsymbol{s}^+ = \boldsymbol{y}_0$ $\boldsymbol{X}\lambda + \sigma \boldsymbol{x}_0 + \boldsymbol{s}^- = \boldsymbol{x}_0$ $\mathbf{1}^T\lambda = 1$ $\lambda, \boldsymbol{s}^-, \boldsymbol{s}^+ \geq \boldsymbol{0}$ $\varepsilon > 0$	*min* $\nu^T \boldsymbol{x}_0 - \mu \boldsymbol{y}_0 + u$ (3.1b) s.t. $-\mu^T\boldsymbol{Y} + \nu^T\boldsymbol{X} + u\mathbf{1}^T \geq \boldsymbol{0}$ $\mu^T \boldsymbol{y}_0 + \nu^T \boldsymbol{x}_0 = 1$ $\mu, \nu \geq \varepsilon \mathbf{1}$ $\varepsilon > 0$

Next we introduce the model when, after the composition of an interval scale variable, the new ratio scale variables are both objectives by character. Assume *t* inputs among the total of *m,* and *s* outputs among the total of *p,* have been measured on the interval scale. Now replace each by two ratio scale variables whose difference is the original variable. The minuend remains as input/output according to what type the original variable was and the subtrahend becomes an output if the original variable was an input and correspondingly the subtrahend of an output variable becomes an input. An example: an output variable profit is replaced by output variable "sum of revenues" and input variable "sum of costs".

Arrange the new set of variables in such a way that the input matrix $\boldsymbol{X} \in \Re_+^{(m+s)xn}$ contains first the t new ratio scale input variables originating from the interval scale input variable (minuends). Next come the s ratio scale variables that originate from the interval scale output variable (subtrahends). As for the output matrix $\boldsymbol{Y} \in \Re_+^{(p+t)xn}$ for convenience we arrange the new output variables originating from the interval scale input variables first (the subtrahends in the difference that corresponds to the interval scale input variable) and next the new output variables corresponding to the original interval scale outputs (minuends).

Consider the efficiency of DMU_0. The coefficients of the new ratio scale variables are set equal in the dual formulation, in other words the shadow prices must coincide. Note that each resulting new constraint in the dual creates a new variable, denoted here by ν, in the primal. Note that we refer to DMU_0 by index '0' except in the vectors $\boldsymbol{X}_i$ and $\boldsymbol{Y}_r$ where its represented by its original subscript.

Figure 3.1 Decomposition of interval scale outputs and inputs

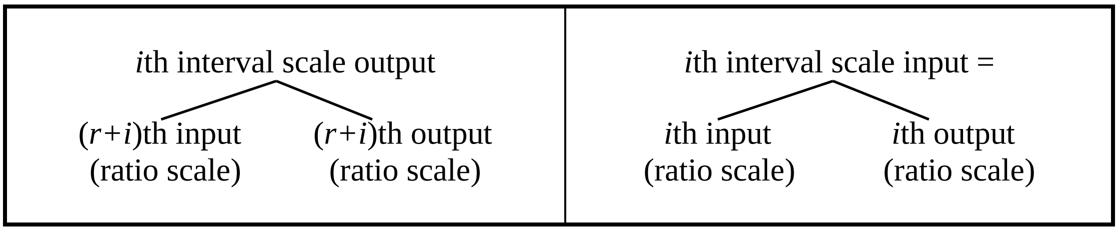

Table 3.2 The combined BCC model with the interval scale variables decomposed into one input and one output each

Combined BCC Primal **(BCC_P - C-R)**	**Combined BCC Dual** **(BCC_D – C-R)**
max $\sigma + \varepsilon \mathbf{1}^T(\boldsymbol{s}^+ + \boldsymbol{s}^-)$ s.t. (3.2a) $\mathbf{Y}_r\lambda - \sigma \boldsymbol{y}_{r0} - \boldsymbol{s}^+_r - \boldsymbol{v}_r = \boldsymbol{y}_{r0}, \quad r=1,\dots,t+s$ $\mathbf{Y}_r\lambda - \sigma \boldsymbol{y}_{r0} - \boldsymbol{s}^+_r = \boldsymbol{y}_{r0}, \quad r=t+s+1,\dots,p+t$ $\mathbf{X}_i\lambda + \sigma \mathbf{x}_{i0} + \boldsymbol{s}^-_i - \boldsymbol{v}_i = \boldsymbol{x}_{i0}, \quad i=1,\dots,t+s$ $\mathbf{X}_i\lambda + \sigma \mathbf{x}_{i0} + \boldsymbol{s}^-_i = \boldsymbol{x}_{i0}, \quad i=t+s+1,\dots,m+s$ $\mathbf{1}^T\lambda = 1$ $\lambda, \mathbf{s}^-, \mathbf{s}^+ \geq \mathbf{0}$ $\varepsilon > 0$ (non-Archimedean)	*Min* $\nu^T \boldsymbol{x}_0 - \mu^T \boldsymbol{y}_0 + u$ s.t (3.2b) $-\mu^T\boldsymbol{Y} + \nu^T\boldsymbol{X} + u\mathbf{1}^T \geq \mathbf{0}$ $\mu^T y_0 + \nu^T x_0 = 1$ $\mu_r - \nu_i = 0, \quad r = 1,\dots,t+s, i=r$ $\mu, \nu \geq \varepsilon\mathbf{1}$ $\varepsilon > 0$ (non-Archimedean)

A DMU is efficient iff the optimal value $\sigma^* = 0$ and all slack variables $\boldsymbol{s}^-, \boldsymbol{s}^+$ equal zero; otherwise it is inefficient (Charnes et al., 1994). Note that in this model the interpretation of σ^* is how much at least each input can be decreased and each output increased compared to an efficient unit.

Naturally apart from the above model, input or output oriented models can be considered. If we set $\boldsymbol{x}_{i0}$, $i = 1,\ldots,m+s,$ to zero in 3.2a) we get the output oriented formulation. The input oriented model is derived analogously.

Setting the new variables' shadow prices equal is not absolutely necessary. That constraint can be omitted. However, intuitively the constraint is appealing – their origin is one variable whose two "sides" they represent. Also, consider model 3.2b and assume one variable, say $z = z^+ - z^-$, has been decomposed into two. Now if the weights of z^+ and z^- are set equal in 3.2b in both the objective and in the first set of constraints actually the **original** variable (multiplied by its weight) appears and only in the normalization row this is not the case.

3.2 The case when one of the new variables is non-discretionary by character

When dealing with an output variable h that is an increment of two observations in time $h = g_{\tau_1} - g_{\tau_2}$ with $\tau 1 > \tau 1$, then the observation at time point *s* is very much like a non-discretionary input variable. Let us think of a bank attracting new accounts or a magazine campaigning for new subscriptions. There is no control over the position where they stand when starting the effort. We write an input-oriented model for that kind of a situation, where the non-discretionary inputs are not among the objectives. In the next formulation in Table 3.2 there are *r* interval scale output variables decomposed into two and the accruing new input variables are non-discretionary by character.

Table 3.3 The input-oriented BCC model with r output variables decomposed into two where the accruing input variables are non-discretionary

Input-oriented BCC Primal (BCC_P - I-R)	Input-oriented BCC Dual (BCC_D - I-R)
max $\sigma + \varepsilon \boldsymbol{1}^T(\boldsymbol{s}^+ + \boldsymbol{s}^-)$ s.t. (3.3a) $\boldsymbol{Y}_r\lambda - \boldsymbol{s}^+_r - \boldsymbol{v}_r = \boldsymbol{y}_{r0}$, $r = 1,\ldots, t$ $\boldsymbol{Y}_r\lambda - \boldsymbol{s}^+_r = \boldsymbol{y}_{r0}$, $r = t+1,\ldots,p$ $\boldsymbol{X}_i\lambda + \boldsymbol{s}^-_i - \boldsymbol{v}_i = \boldsymbol{x}_{i0}$, $i = 1 \ldots,t$ $\boldsymbol{X}_i\lambda + \sigma \boldsymbol{x}_{i0} + \boldsymbol{s}^-_i = \boldsymbol{x}_{i0}$, $i = t+1,\ldots,m+t$ $\boldsymbol{1}^T\lambda = 1$ $\lambda,\ \boldsymbol{s}^-, \boldsymbol{s}^+ \geq \boldsymbol{0}$ $\varepsilon > 0$	*Min* $\nu^T \boldsymbol{x}_0 - \mu^T \boldsymbol{y}_0 + u$ s.t (3.3b) $-\mu^T\boldsymbol{Y} + \nu^T\boldsymbol{X} + u\boldsymbol{1}^T \geq \boldsymbol{0}$ $\nu^T\boldsymbol{x}_o = 1$ $\mu_r - \nu_i = 0$, $r = 1 ,\ldots, t+s,\ i=r$ $\mu, \nu \geq \varepsilon \boldsymbol{1}$ $\varepsilon > 0$

3.3 The efficiency/inefficiency of the units after the analysis with decomposed variables

If a unit is efficient, it is necessary that it maintains this status even if the set of inputs and outputs is modified in the way proposed in this paper. We introduce the following definition

Definition 3.1 Unit 0 with an input vector $\boldsymbol{x}_0 \in \Re^m$ and an output vector $\boldsymbol{y}_0 \in \Re^p$ is efficient with respect to set S if there does not exist a vector of coefficients $\boldsymbol{a} \in S \subseteq \Re_+^n$ such that

$$\boldsymbol{Y}_r\boldsymbol{a} \geq y_{r0}\,,\ r=1,\ldots,p$$
$$\boldsymbol{X}_i\boldsymbol{a} \leq x_{i0}\ ,\ i=1,\ldots m$$

with at least one strict inequality. Note that even negative values for the input and output vectors are allowed.

If $S = \{\boldsymbol{a} / a_i \in \{0, 1\}, \boldsymbol{1}^T \boldsymbol{a} = 1\}$, then Pareto efficiency (or nondominance) is in question That means that if a unit is not efficient there is a superior unit in the set of units under consideration.

It is easy to see that unit 0 with an input vector $\boldsymbol{x}_0 \in \Re_+^m$ and an output vector $\boldsymbol{y}_0 \in \Re_+^p$ is efficient with respect to set $S = \{\boldsymbol{a} \geq \boldsymbol{0} \,/\, \boldsymbol{1}^T \boldsymbol{a} = 1\}$ iff in the optimum of 3.1a $\boldsymbol{\sigma^*=0}$, $\boldsymbol{s}^-{}^*=\boldsymbol{0}$ and $\boldsymbol{s}^+{}^*=\boldsymbol{0}$. The result is true also for $S = \{\boldsymbol{a} \geq \boldsymbol{0}\}$ and 3.1a without the restriction $\boldsymbol{1}^T\lambda = 1$.

Next we prove an efficient unit cannot become inefficient in model 3.2a after the decomposition of one output variable, originally possibly negative. Without loss of generality we assume that output has index p. The proof for more outputs and inputs is straightforward as well as the proof for the CCR model.

Theorem 3.1
Consider unit 0 with an input vector $\boldsymbol{x}_0 \in \Re_+^m$ and outputs $y_{r0} \in \Re_+$, r=1,…,p-1, $y_{p0} \in \Re$. Assume that the unit is efficient in the sense that there does not exist any positive vector of coefficients $\boldsymbol{a} \in S = \{\boldsymbol{a} \geq \boldsymbol{0} \,/\, \boldsymbol{1}^T \boldsymbol{a} = 1\}$ such that

$$\boldsymbol{Y}_r\boldsymbol{a} \geq y_{r0}\,,\ r=1,\ldots,p$$
$$\boldsymbol{X}_i\boldsymbol{a} \leq x_{i0}\ ,\ i=1,\ldots m$$

with at least one strict inequality. Then after decomposing y_p, so that $y_p = z - u$, where z is to be maximized (output) and u is to be minimized (input) and diagnosing this modified set of inputs and outputs by 3.2a, unit 0 is diagnosed efficient .

Proof

Assume that the optimal solution of 3.2a is $\boldsymbol{\lambda^*}, \boldsymbol{\sigma^*}$, $\boldsymbol{s}^{-*}$, $\boldsymbol{s}^{+*}$, v^* and that unit 0 is diagnosed inefficient. Then it is necessary that at least one from the following list of variables is strictly positive: $\boldsymbol{\sigma^*}$, an element of $\boldsymbol{s}^{-*}$, an element of $\boldsymbol{s}^{+*}$. That implies

$$\boldsymbol{Y}_r\boldsymbol{\lambda^*} \geq y_{r0}, \; r=1,\ldots,p-1 \qquad (3.4a)$$
$$\boldsymbol{X}_i\boldsymbol{\lambda^*} \leq x_{i0}, \; i=1,\ldots m. \qquad (3.4b)$$

and

$$\boldsymbol{z}^T\boldsymbol{\lambda^*} - v^* \geq z_0 \qquad (3.4c)$$
$$\boldsymbol{u}^T\boldsymbol{\lambda^*} - v^* \leq u_0 \qquad (3.4d)$$

with at least one strict inequality. Rewriting (3.4c-d) we get

$$(\boldsymbol{z} - \boldsymbol{u})\boldsymbol{\lambda^*} - v^* + v^* \geq z_0 - u_0 \text{ and}$$
$$\boldsymbol{Y}_p\,\boldsymbol{\lambda^*} \geq y_{p0}\,. \qquad (3.4e)$$

There cannot exist any $\boldsymbol{\lambda^*}$ with at least one strict inequality in 3.4a-b or 3.4e because that is in conflict with the assumption that the original input-output vector was efficient. Q.E.D.

Note that even if we did not impose any restrictions for the weights in 3.2b, efficient units remain efficient after the decomposition. The increase of variables in DEA means, however, also in this case that inefficient units may become efficient.

4. Other Approaches Dealing with Interval Scale Data

4.1 CCR and BCC models with general weights

As frequently mentioned, for an interval scale variable division is not allowed, only differences of variable values can be calculated. Thus it is possible to calculate the difference of a “good” (efficient) value and another observation of the variable. To be able to compare a difference with other input/output differences having different scales the differences must be considered relative to something scaling the differences. In a radial model the something is the value of the inputs and outputs of the unit under consideration. That is only one choice, we generally speaking need **some jointly elected measure**. In his study Korhonen (1997) has selected the **range** of the variable in the data to be this scaling measure and the observed values (of the inputs and outputs) are considered with respect to the range throughout.

Other good measures may exist. The Decision Maker finally utilizing the results of the analysis may have an idea what the measure is. When studying the results of the analysis she/he then is aware of the basis the calculations.

When the model is not radial, the coefficients of σ in 3.1a are replaced by other positive values, such as range. In table 3.2 we denote these subjectively chosen strictly positive vectors of weights by $\boldsymbol{w}^x$ and $\boldsymbol{w}^y$ which replace $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ in the radial model as coefficients of σ. Now the projection of an inefficient unit on the efficient frontier can be any (virtual) unit dominating it. No translation or modifications in the original data are needed.

Table 4.1 Combined Nonradial BCC Model

Nonradial Combined BCC Primal (BCC_P - C-NR)	**Nonradial Combined BCC Dual (BCC_D - C-NR)**
max $\sigma + \varepsilon \boldsymbol{1}^T(\boldsymbol{s}^+ + \boldsymbol{s}^-)$ (4.1a) s.t $\boldsymbol{Y}\lambda - \sigma \boldsymbol{w}^y - \boldsymbol{s}^+ = \boldsymbol{y}_0$ $\boldsymbol{X}\lambda + \sigma \boldsymbol{w}^x + \boldsymbol{s}^- = \boldsymbol{x}_0$ $\boldsymbol{1}^T\lambda = 1$ $\lambda, \boldsymbol{s}^-, \boldsymbol{s}^+ \geq \boldsymbol{0}$ $\varepsilon > 0$	*min* $\nu^T \boldsymbol{x}_0 - \mu \boldsymbol{y}_0 + u$ (4.1b) s.t. $-\mu^T\boldsymbol{Y} + \nu^T\boldsymbol{X} + u\boldsymbol{1}^T \geq \boldsymbol{0}$ $\mu^T \boldsymbol{w}^y + \nu^T\boldsymbol{w}^x = 1$ $\mu, \nu \geq \varepsilon \boldsymbol{1}$ $\varepsilon > 0$

Note that $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ need not be positive. The approach maintains the efficiency/inefficiency status of units (each input and output observation is divided by its range in the data) and only the scores of the inefficient units change.

4.2 Models that do not produce efficiency scores

Ali and Seiford (1990) discovered that the additive model (Charnes et al. 1985) is translation invariant. That model does not, however, produce efficiency scores which, we feel, is a serious drawback. Lovell and Pastor (1995) and Pastor (1994) suggest the weighted additive model, which is able to produce an "efficiency index" in the case where the "variables are prices (or scaled) in such a way they stay on an absolutely equal footing". This seems, however, to be rarely the case. If something speaks strongly for the CCR and BCC models and efforts for their further development, it is their familiarity.

Zhu (1994) also considers negative inputs and outputs. He proposes a somewhat complicated procedure for calculating efficiency scores translating the data and using ideas of controlled envelopment analysis.

5. Example

Consider the efficiency of the forwards of the ice-hockey team Porin Ässät which plays in the Finnish Ice Hockey League. In ice-hockey an important measure of performance for a player is the figure "goals scored by the *own team* minus goals scored by the *opponent team* while the player is on ice". No powerplay goals are taken into account in the figure. In the data in Table 5.1 they are simply called the Plusgoals and the Minusgoals and the result of their deduction is Points +/-. The last figure is clearly measured on an interval scale and is frequently negative. Another output variable we use is the points achieved by the player which consist of goals scored and assisted. As an input variable we use games played. The information on the minutes played by a player is not available but according to ice-hockey experts the players have relatively evenly time on the ice in the Finnish Ice Hockey League.

Table 5.1 The players and their input output variable values (season 1996/97)

Players	Games	Points	Plusgoals	Minusgoals	points +/-
Vujtek	50	58	38	54	-16
Fandul	48	48	32	49	-17
Korpisalo	49	45	33	39	-6
Poulsen	49	35	36	31	5
Virta	48	34	35	31	4
Levonen	50	25	19	31	-12
Mikkola	47	25	31	28	3
Alinc	47	25	26	40	-14
Salonen	45	16	30	28	2
Saarinen	48	15	31	35	-4
Kotkaniemi	47	14	21	20	1
Virtanen	35	13	19	18	1
Karapuu	19	4	7	4	3
Tuominen	23	4	6	11	-5

Both the variables goals scored and goals scored by the opponent team are objectives. The goals scored by the opponent team can be viewed either as outputs or inputs (to be minimized). We considered them as inputs. Scores produced by several models are calculated. We calculated also the scores with generalized weights (ranges) as discussed in 4.1. In the following Vujtek's problems (5.1-5.2) are introduced. Model 5.1 is the combined BCC model and 5.2 is the combined nonradial BCC model with ranges.

$$max\ \sigma + \varepsilon \mathbf{1}^T(\mathbf{s}^+ + \mathbf{s}^-)$$

$$\text{s.t.} \quad \begin{aligned} &38\lambda_1 + 32\lambda_2 + 33\lambda_3 + ... + 6\lambda_{14} - \sigma 38 - s^+_1 - v_1 = 38 \\ &58\lambda_1 + 48\lambda_2 + 45\lambda_3 + ... + 4\lambda_{14} - \sigma 58 - s^+_2 = 58 \\ &54\lambda_1 + 49\lambda_2 + 39\lambda_3 + ... + 11\lambda_{14} + \sigma 54 + s^-_1 - v_1 = 54 \\ &50\lambda_1 + 48\lambda_2 + 49\lambda_3 + ... + 23\lambda_{14} + \sigma 50 + s^-_2 = 50 \\ \\ &\mathbf{1}^T\boldsymbol{\lambda} = 1 \\ &\boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^- \geq \mathbf{0} \\ &\varepsilon > 0 \end{aligned} \qquad (5.1)$$

$$max\ \sigma + \varepsilon \mathbf{1}^T(\mathbf{s}^+ + \mathbf{s}^-)$$

$$\text{s.t.} \quad \begin{aligned} &-16\lambda_1 -17\lambda_2 - 6\lambda_3 + ... - 5\lambda_{14} - \sigma 22 - s^+_1 = -16 \\ &58\lambda_1 + 48\lambda_2 + 45\lambda_3 + ... + 4\lambda_{14} - \sigma 54 - s^+_2 = 58 \\ &50\lambda_1 + 48\lambda_2 + 49\lambda_3 + ... + 23\lambda_{14} + \sigma 31 + s^-_1 = 50 \\ \\ &\mathbf{1}^T\boldsymbol{\lambda} = 1 \\ &\boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^- \geq \mathbf{0} \\ &\varepsilon > 0 \end{aligned} \qquad (5.2)$$

The calculation of the scores can be done with any LP solver.

In the following tables improvement needs for the forwards according to different DEA models are presented. From the oriented models we calculated only the output oriented one because it seemed to fit better for the problem.

Table 5.2 The improvement needs for Porin Assat forwards according to radial models

MODELS BASED ON RADIAL MEASUREMENTS OF EFFICIENCY				
	BCC combined	BCC output oriented	CCR combined	CCR output oriented
Vujtek	0,0000	0,0000	0,0000	0,0000
Fandul	0,0477	0,1086	0,0544	0,1151
Korpisalo	0,0084	0,0185	0,0100	0,0200
Poulsen	0,0000	0,0000	0,0000	0,0000
Virta	0,0060	0,0133	0,0086	0,0174
Levonen	0,2179	0,6247	0,2420	0,6385
Mikkola	0,0300	0,0602	0,0391	0,0813
Alinc	0,1895	0,5298	0,2132	0,5418
Salonen	0,0448	0,0911	0,0645	0,1379
Saarinen	0,1291	0,2882	0,1445	0,3379
Kotkaniemi	0,0876	0,1841	0,1168	0,2645
Virtanen	0,0780	0,1614	0,0871	0,1909
Karapuu	0,0000	0,0000	0,0000	0,0000
Tuominen	0,1581	1,2434	0,4037	1,3540

Table 5.3 The improvement needs for Porin Assat forwards according to models with improvements relative to the range

MODELS WITH IMPROVEMENTS RELATIVE TO THE RANGE				
	BCC combined	BCC output oriented	CCR combined	CCR output oriented
Vujtek	0,0000	0,0000	0,0000	0,0000
Fandul	0,0603	0,1149	0,0854	0,1219
Korpisalo	0,0132	0,0191	0,0157	0,0209
Poulsen	0,0000	0,0000	0,0000	0,0000
Virta	0,0085	0,0122	0,0120	0,0160
Levonen	0,2594	0,3665	0,2801	0,3726
Mikkola	0,0776	0,0848	0,0817	0,0981
Alinc	0,2484	0,3588	0,2767	0,3680
Salonen	0,1136	0,1242	0,1304	0,1566
Saarinen	0,2545	0,3676	0,2792	0,3714
Kotkaniemi	0,1607	0,1758	0,1775	0,2132
Virtanen	0,0966	0,1394	0,1161	0,1394
Karapuu	0,0000	0,0000	0,0000	0,0000
Tuominen	0,0645	0,1290	0,1965	0,2614

A player is efficient if the score is 0. Figure 0.11 means the player can improve his performance by 11 per cent. In the radial model the 11 per cent means the improvement needs relative to the player's own input and output values and in the range model it means 11 per cent of the range of each variable (in the output-oriented models the improvement needs concern only outputs).

As can be seen for a score it is relevant if we use the combined or the ouput-oriented model. The CCR and BCC models do not make much difference in this data because the players' figures are not very different from each other and in the optimum of the CCR model calculations the sum of the optimal weights of the primal for each unit is very close to 1. Also the range model gives very much the same results in this case.

If units are only diagnosed efficient/inefficient (without any score produced) then, in the case of the BCC and CCR models, we can read the results from Table 5.3 . If Pareto efficient (nondominated) vectors are searched we may state that Vujtek, Fandul, Korpisalo, Poulsen, Mikkola, Salonen, Virtanen and Karapuu are efficient and the rest of the players are inefficient.

6. Summary and Conclusions

The approach introduced in Section 3 allows us to use radial models for interval scale variables. The CCR models are not translation invariant and thus the procedures we propose are the only ones we know that could be used for them. In the BCC models some translation invariance properties are available but they restrict the orientation of the model.

We find the problem of interval scale variables important. In the radial procedure proposed, though the number of variables increases, each of them is interpretable. Efficient units remain efficient. The proposed procedure can be useful in practical applications especially when the constant returns to scale assumption is valid.

We proposed that in addition to the decomposition the weights of the new pair of variables could be set equal. This is not, however, absolutely necessary.

Both the radial model proposed and the model with generalized weights are valid approaches for the interval scale data problem. When to use which then? The radial model and the generalized weights model both maintain efficient units efficient. The advantage of the radial model is that as the new variables are interpretable then it comes very close to the classical familiar DEA models. As for the generalized weight model the main advantage is that the data can be used as such, no modifications are needed. In that model, however, the analyst and the end-user have to agree on what measure to use.

References

Ali, A.I. and Seiford, L.M. (1990) “Translation Invariance in Data Envelopment Analysis”, **Operations Research Letters** 9, 403-405.

Banker , R.D. and Morey, R.C. (1986), “Efficiency Analysis for Exogenously Fixed Inputs and Outputs”, **Operations Research**, 34:4, 513-521.

Chambers, R.G., Chung, Y. and Färe, R.(1996) “Benefit and Distance Functions”, **Journal of Economic Theory**, 70:2.

Charnes, A., Cooper,W.W., Golany, B., Seiford, L. and Stutz, J. (1985) Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production **Functions. Journal of Econometrics** (Netherlands). 30(1/2), 91-107.

Charnes, A., Cooper,W.W. and Rhodes, E. (1978) “Measuring Efficiency of Decision Making Units”, **European Journal of Operational Research** 2, 429-444.

Deprins, D., Simar, L. and Tulkens, H.(1984) “Measuring Labor-Efficiency in Post Offices”, in Marchand, M., Pestieau, P. and Tulkens, H. **The Performance of Public Enterprises: Concepts and Measurements**, Amsterdam: North-Holland.

Halme, M., Joro, T, Korhonen, P., Salo, S. and Wallenius, J. (1998) “A Value Efficiency Approach to Incorporating Preference Information in Data Envelopment Analysis, forthcoming in **Management Science.**

Joro, T., Korhonen, P. and Wallenius, J (1998), “Structural Comparison of Data Envelopment Analysis and Multiple Objective Linear Programming, **Management Science** 44:7, 962-970

Korhonen, P. (1997) personal communication.

Lovell, C.A.K., (1995) “Measuring the Macroeconomic Performance of the Taiwanese Economy”, **International Journal of Production Economics** 39, 165-178.

Lovell, C.A.K. and Pastor, J. (1995) “Units Invariant and Translation Invariant DEA Models”, **Operations Research Letters** 18, 147-151.

Pastor, J. (1994) “New Additive DEA Models for Handling Zero and Negative Data”, unpublished manuscript.

Pastor, J. (1996) “Translation Invariance in Data Envelopment Analysis: A Generalization”, **Annals of Operation Research** 66, 93-102.

Thanassoulis, E., Boussofiane, R. and Dyson, R.G. (1995) “Exploring output quality targets in the provision of perinatal care in England usind Data Envelopment Analysis”, **European Journal of Operational Research**, 80, 588-607.

Thompson. R.G., Langemeier, L.N., Lee, Chi-Tah, Lee, Euntaik, Thrall, R.M. (1990) “The Role of Multiplier Bounds in Efficiency Analysis with Application to Kansas Farming”, **Journal of Econometrics** 46, 93-108.

Thompson, R.G., Singleton, F., Thrall, R. and Smith, B. (1986) “Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas”, **Interfaces**, 16, 35-49.

Zhu, J. (1994) “Translation Invariance in Data Envelopment Analysis and Its Use”, unpublished manuscript.