



Munich Personal RePEc Archive

A universal solution for units-invariance in data envelopment analysis

Xu, Jin and Zervopoulos, Panagiotis and Qian, Zhenhua and
Cheng, Gang

29 September 2012

Online at <https://mpra.ub.uni-muenchen.de/41633/>
MPRA Paper No. 41633, posted 01 Oct 2012 13:35 UTC

A universal solution for units-invariance in data envelopment analysis

Jin Xu

China Center for Health Development Studies, Peking University, 38 Xueyuan Rd,
Beijing, China

xujin@bjmu.edu.cn

Panagiotis D. Zervopoulos

China Center for Health Development Studies, Peking University, 38 Xueyuan Rd,
Beijing, China

Department of Business Administration of Food and Agricultural Enterprises,
University of Ioannina, 2 Georgiou Seferi St, Agrinio, Greece

pzervopoulos@uoi.gr

Zhenhua Qian

School of Social Science, University of Science and Technology Beijing, 30 Xueyuan Rd,
Beijing, China

rosezhenhua@yahoo.com.cn

Gang Cheng*

China Center for Health Development Studies, Peking University, 38 Xueyuan Rd,
Beijing, China

chenggang@bjmu.edu.cn

*Corresponding author

Abstract: The directional distance function model is a generalization of the radial model in data envelopment analysis (DEA). The directional distance function model is appropriate for dealing with cases where undesirable outputs exist. However, it is not a units-invariant measure of efficiency, which limits its accuracy. In this paper, we develop a data normalization method for DEA, which is a universal solution for the problem of units-invariance in DEA. The efficiency scores remain unchanged when the original data are replaced with the normalized data in the existing units-invariant DEA models, including the radial and slack-based measure models, i.e., the data normalization method is compatible with the radial and slack-based measure models. Based on normalized data, a units-invariant efficiency measure for the directional distance function model is defined.

Keywords: Data Envelopment Analysis; Data normalization; Units-invariance; Directional distance function

1. Introduction

Data Envelopment Analysis (DEA), originally developed by Charnes et al. (1978), is a nonparametric method that draws on linear programming for measuring the comparative efficiency of Decision making Units (DMUs). DEA has been applied extensively in many different areas (Cook & Seiford, 2009; Seiford, 1996). A fundamental advantage of DEA is units-invariance,

which means that efficiency scores assigned to DMUs are independent of the measurement units of the inputs and outputs that are utilized in the assessment process (Lovell & Pastor, 1995; Tone, 2001). Radial DEA models, such as CCR and BBC models (Banker et al., 1984; Charnes, 1994), and the radial measure, such as the slack-based measure (SBM), are units-invariant (Färe & Knox Lovell, 1978; Tone, 2001).

The directional distance function model is a generalization of radial models (Chambers et al., 1996; Chambers et al., 1998; Chung et al., 1997). A special feature of the directional distance function model is that the direction the DMUs under evaluation are projected to the production frontier can be customized. By assigning a direction vector in Euclidean space, one can project the evaluated DMU on a specific point on the frontier. Particularly when the direction vector points towards the origin of the coordinates, the directional distance function model is equivalent to the radial model. Two advantages of the directional distance function model are that: 1) researchers can specify the direction of decreasing inputs and increasing outputs by assigning a direction vector, and 2) researchers can easily deal with the cases where undesirable outputs exist. However, a drawback of the directional distance function is that its measurement is generally not units-invariant. Taking into account that the inputs and outputs of the evaluated DMUs serve as the direction vector, changes in the measurement units of inputs or outputs potentially can lead to significant differences in the results.

The proposed data normalization method provides a universal solution when the applied DEA model violates the units-invariance criterion. The properties of the proposed method are tested with the DEA-based directional distance function model, but the new method can be applied to all existing and future DEA models.

2. The method of data normalization of DEA and its properties

The measurement of efficiency using radial DEA models is not affected by the measurement units of inputs and outputs because efficiency results from the comparison of the inputs and outputs of the evaluated DMU against the corresponding values of the target DMU (benchmark). For radial models, the inputs or outputs are improved in proportion. In non-radial models, such as SBM models, the “proportional improvement” restriction is loosened, but the measurement of efficiency still draws on input-output ratios. As a result, efficiency scores are not affected by the measurement units of inputs and outputs.

The concept used to develop a method for dealing with the issue of units-invariance is based on the introduction of a preparation stage prior to the application of DEA models. In this stage, the original input and output data are converted into dimensionless data. When dimensionless data are utilized, this stage ensures that the efficiency scores produced by any DEA model will meet the units-invariance criterion.

The proposed procedure is expected to satisfy the conditions below:

- (1) The data conversion should not affect efficiency scores measured by any units-invariant radial or non-radial DEA model.

- (2) The results produced by DEA models using converted data can be converted reversely so that to be completely consistent with the results obtained from DEA models utilizing original data. The consistency of the results should be expected regardless of the units-invariance DEA model (i.e., radial or non-radial) that is applied.

The above two conditions ensure the proposed model's compatibility with existing units-invariant DEA measures.

Taking into account the points raised above, in this paper, we develop a DEA data normalization method.

Let m represents the number of inputs and q represents the number of outputs for each of the n DMUs. Column vectors x_j and y_j express the inputs and outputs, respectively, of DMU _{j} , \hat{x}_j and \hat{y}_j denote the normalized value of inputs and outputs, respectively; and x_0 and y_0 stand for the original inputs and outputs, respectively, of the evaluated DMU (DMU₀). A conversion is applied as follows

$$\begin{aligned} \hat{x}_{ij} &= x_{ij} / x_{i0}, \quad i = 1, 2, \dots, m \\ \hat{y}_{rj} &= y_{rj} / y_{r0}, \quad r = 1, 2, \dots, q \\ j &= 1, 2, \dots, n \end{aligned} \quad (1)$$

The normalization formula can be extended to inputs or outputs with negative values as follows

$$\begin{aligned} \hat{x}_{ij} &= x_{ij} / |x_{i0}|, \quad i = 1, 2, \dots, m \\ \hat{y}_{rj} &= y_{rj} / |y_{r0}|, \quad r = 1, 2, \dots, q \\ j &= 1, 2, \dots, n \end{aligned} \quad (2)$$

Essentially, the inputs (outputs) of DMU₀ serve as measurement units for every input (output) of the sample. The data conversion presented in formulas (1) and (2) does not affect the efficiency scores measured by any DEA model that is originally units-invariant.

Unlike other data normalization methods, the proposed data normalization for DEA yields one discrete normalized dataset for each DMU _{j} , i.e., there will be n normalized data sets for the n DMUs of the sample.

Data normalization for DEA has the following properties:

- (1) DEA data normalization is a dimension-free conversion. Regardless of the measurement units of the original inputs and outputs or even the changes in the measurement units used with the original inputs and outputs, the normalized data remain the same.
- (2) All the inputs and outputs of DMU₀ are equal to unity after normalization.

Subsequent to data normalization, the DEA models that are originally units-invariant yield efficiency scores that are identical to those obtained when non-normalized data are used. In addition, when normalized data are used, the slacks generated from DEA models can be converted reversely, as follows

$$s_i^- = \hat{s}_i^- \times x_{i0}, \quad s_r^+ = \hat{s}_r^+ \times y_{r0} \quad (3)$$

where s stands for reversely converted slacks, \hat{s} are the slacks identified by the DEA model when normalized data are utilized, x_{i0} and y_{r0} express the inputs and outputs, respectively, of DMU₀.

The input-oriented CRS model using raw data can be expressed as

$$\begin{aligned} \min \quad & \theta \\ \text{s.t.} \quad & X\lambda + s^- - \theta x_0 = 0 \\ & Y\lambda - s^+ = y_0 \\ & \lambda, s^-, s^+ \geq 0 \end{aligned} \quad (4)$$

The output-oriented CRS model using raw data can be expressed as:

$$\begin{aligned} \max \quad & \varphi \\ \text{s.t.} \quad & X\lambda + s^- = x_0 \\ & Y\lambda - s^+ - \varphi y_0 = 0 \\ & \lambda, s^-, s^+ \geq 0 \end{aligned} \quad (5)$$

In radial DEA models, radial movement and slack movement are negative for inputs, and positive for outputs. The relationship between the original inputs (outputs), radial movements, slack movements, and target inputs (outputs) are formulated as follows

Target value = original value + radial movement + slack movement

$$X\lambda = x_0 + (\theta - 1)x_0 + (-s^-) \quad (6)$$

$$Y\lambda = y_0 + (\varphi - 1)y_0 + s^+ \quad (7)$$

where $(\theta - 1)$ expresses the radial movement of the input in model (6), and $(\varphi - 1)$ denotes the radial movement of the output in model (7).

After normalization of the data, the input-oriented CRS model becomes

$$\min \quad \theta$$

$$\begin{aligned}
\text{s.t. } & \hat{X}\lambda + \hat{s}^- - \theta\hat{x}_0 = 0 \\
& \hat{Y}\lambda - \hat{s}^+ = \hat{y}_0 \\
& \lambda, \hat{s}^-, \hat{s}^+ \geq 0 \quad (8)
\end{aligned}$$

Respectively, after normalization of the data, the output-oriented CRS model is written as

$$\begin{aligned}
\max \quad & \varphi \\
\text{s.t. } & \hat{X}\lambda + \hat{s}^- = \hat{x}_0 \\
& \hat{Y}\lambda - \hat{s}^+ - \varphi\hat{y}_0 = 0 \\
& \lambda, \hat{s}^-, \hat{s}^+ \geq 0 \quad (9)
\end{aligned}$$

According to property (2) of the data normalization method for DEA, when normalized data are utilized in radial DEA model, all of the inputs and outputs of DMU_o are equal to unity. As a result, formulas (6) and (7) can be rewritten as

$$\hat{X}\lambda = 1 + (\theta - 1) + (-\hat{s}^-) \quad (10)$$

$$\hat{Y}\lambda = 1 + (\varphi - 1) + \hat{s}^+ \quad (11)$$

The non-oriented CRS-SBM model can be expressed as

$$\begin{aligned}
\min \quad \theta &= \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{s_i^-}{x_{i0}}}{1 + \frac{1}{q} \sum_{r=1}^q \frac{s_r^+}{y_{r0}}} \\
\text{s.t. } & X\lambda + s^- = x_0 \\
& Y\lambda - s^+ = y_0 \\
& \lambda, s^-, s^+ \geq 0 \quad (12)
\end{aligned}$$

After normalization of the data, model (12) becomes

$$\begin{aligned}
\min \quad \theta &= \frac{1 - \frac{1}{m} \sum_{i=1}^m \hat{s}_i^-}{1 + \frac{1}{q} \sum_{r=1}^q \hat{s}_r^+} \\
\text{s.t. } & \hat{X}\lambda + \hat{s}^- = \hat{x}_0
\end{aligned}$$

$$\hat{Y}\lambda - \hat{s}^+ = \hat{y}_0$$

$$\lambda, \hat{s}^-, \hat{s}^+ \geq 0 \quad (13)$$

In model (13), the inefficiency is expressed as the average of the slacks identified when normalized data are applied.

In order to prove empirically the consistency of the efficiency scores when normalized data are incorporated in units-invariant DEA models, we refer to Table 1. The testing sample consists of seven DMUs with two inputs (x_1 and x_2) and one output (y). Let DMU G be the unit under evaluation (DMU₀) and apply the input-oriented CRS model to original (raw) and normalized data. The normalized data illustrated in Table 1 are calculated using formula (1). The efficiency score obtained from raw data is identical with the score that resulted from the utilization of normalized data. In a radial model, radial movement (-0.31) represents the degree of inefficiency.

Table 1. Illustration of DEA data normalization: efficiency measurement of DMU G using the input-oriented CRS model

| DMU | Raw data | | | Normalized data | | |
|------------------|----------|--------|-------|-----------------|-------------|-----------|
| | x_1 | x_2 | y | \hat{x}_1 | \hat{x}_2 | \hat{y} |
| A | 10.00 | 40.00 | 10.00 | 0.20 | 0.67 | 0.50 |
| B | 15.00 | 25.00 | 10.00 | 0.30 | 0.42 | 0.50 |
| C | 32.00 | 24.00 | 16.00 | 0.64 | 0.40 | 0.80 |
| D | 48.00 | 16.00 | 16.00 | 0.96 | 0.27 | 0.80 |
| E | 24.00 | 48.00 | 16.00 | 0.48 | 0.80 | 0.80 |
| F | 54.00 | 27.00 | 18.00 | 1.08 | 0.45 | 0.90 |
| G | 50.00 | 60.00 | 20.00 | 1.00 | 1.00 | 1.00 |
| Efficiency score | 0.69 | | | 0.69 | | |
| Radial movement | -15.62 | -18.75 | -0.00 | -0.31 | -0.31 | -0.00 |
| Slack movement | -0.00 | -0.00 | -0.00 | -0.00 | -0.00 | 0.00 |
| Projection | 34.38 | 41.25 | 20.00 | 0.69 | 0.69 | 1.00 |

3. Efficiency measurement using the directional distance function model

The linear programming of the directional distance function model is defined as follows

$$\max \beta$$

$$\text{s.t. } X\lambda + \beta v \leq x_0$$

$$Y\lambda - \beta u \geq y_0$$

$$\lambda, v, u \geq 0 \quad (14)$$

where v and u denote the input and output direction vectors, respectively.

In directional distance function models, direction vectors determine the directions of movement of the inputs and outputs of the inefficient DMUs and target values (projections on the frontier), thereby determining efficiency scores. Direction vectors also reflect the relative importance of inputs and outputs in efficiency measurement. Figure 1 illustrates the impact of direction vectors on efficiency measurement drawing on an input-oriented CRS directional distance function model using normalized data.

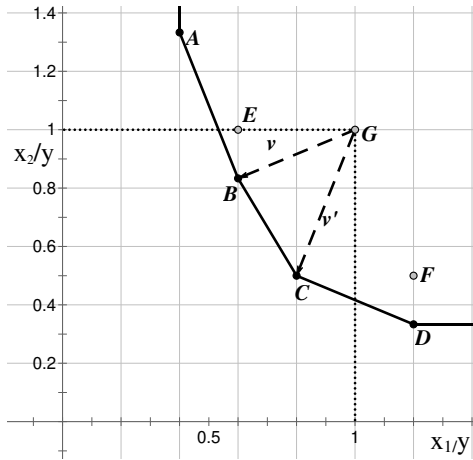


Figure 1. An input-oriented CRS directional distance function model using normalized data

In Figure 1, the horizontal coordinate represents the consumption of x_1 for each unit of output and the vertical coordinate represents the consumption of x_2 for each unit of output. When the direction vector is parallel to the horizontal axis, i.e., $v = (1, 0)$, improvement is applied solely to x_1 , and the efficiency score is determined exclusively by the inefficiency of x_1 . Similarly, when direction vector is parallel to the vertical axis, i.e., $v = (0, 1)$, improvement is associated only with x_2 , and the efficiency score is determined exclusively by the inefficiency of x_2 . Furthermore, a downward movement of the direction vector, i.e., from v to v' , indicates a decrease of the impact of x_1 on the measurement of the efficiency score and an increase of the impact of x_2 .

When the directional distance function models are incorporated in DEA, the inputs and outputs of DMU_0 usually are utilized as direction vectors. In such situations, directional distance function models are equivalent to radial DEA models, and β , which reflects the degree of inefficiency, has the property of units-invariance. Unless the direction vectors are equal to the inputs and outputs of the DMU under evaluation, β is no longer units-invariant. Previous studies have not developed a solution for the problem of units-variance. As a result, the applicability of directional distance function models in efficiency measurement is limited.

Drawing on the definition of SBM model, we defined a units-invariant directional distance function model on the basis of DEA data normalization as follows

$$\theta = \frac{1 - \frac{1}{m} \sum_{i=1}^m \beta v_i}{1 + \frac{1}{q} \sum_{r=1}^q \beta u_r} = \frac{1 - \beta \frac{1}{m} \sum_{i=1}^m v_i}{1 + \beta \frac{1}{q} \sum_{r=1}^q u_r}$$

$$\max \quad \beta$$

$$\text{s.t.} \quad \hat{X} \lambda + \beta v \leq \hat{x}_0$$

$$\hat{Y} \lambda - \beta u \geq \hat{y}_0$$

$$\lambda, v, u \geq 0 \quad (15)$$

where βv and βu represent the inefficiency of the inputs and outputs, respectively. The inefficiency score of the evaluated DMU is calculated as the arithmetical mean of the inefficiency scores of inputs and outputs.

In model (15), when the input direction vector v is set equal to the input vector of DMU_o, i.e., $v = (1, 1, \dots, 1)$, and the output direction vector u is assigned a null vector, the model becomes equivalent to the input-oriented radial DEA model using normalized data, with efficiency score $\theta = 1 - \beta$. The efficiency score obtained from the application of model (15) is identical with the results obtained from radial models (4) and (8).

Alternatively, in model (15), by assigning a null vector to the input direction vector v , and setting the output direction vector u equal to the output vector of DMU_o, i.e., $u = (1, 1, \dots, 1)$, the directional distance function model becomes equivalent to the output-oriented DEA model using normalized data. In this case, the efficiency score is defined as $\theta = 1/(1 + \beta)$. The efficiency score calculated by the directional distance function model (15) is identical with the results provided by radial models (5) and (9).

Theorem 1: For the data set illustrated in Table 1, if the length of the direction vector changes and the direction of the same vector is unchanged, then the efficiency remains unchanged.

Proof: Assume that the direction vectors of input and output are scaled up proportionally from v and u to bv and bu , respectively, with b being a positive real number, and the Euclidian directions of the vectors are unchanged. Thus, model (15) becomes

$$\theta = \frac{1 - \beta b \frac{1}{m} \sum_{i=1}^m v_i}{1 + \beta b \frac{1}{q} \sum_{r=1}^q u_r}$$

$$\max \quad \beta b$$

$$\text{s.t.} \quad \hat{X} \lambda + \beta b v \leq \hat{x}_0$$

$$\hat{Y} \lambda - \beta b u \geq \hat{y}_0$$

$$\lambda, \nu, u \geq 0 \quad (16)$$

If we let α be equal to βb

$$\theta = \frac{1 - \alpha \frac{1}{m} \sum_{i=1}^m v_i}{1 + \alpha \frac{1}{q} \sum_{r=1}^q u_r}$$

$$\max \quad \alpha$$

$$\text{s.t.} \quad \hat{X} \lambda + \alpha \nu \leq \hat{x}_0$$

$$\hat{Y} \lambda - \alpha u \geq \hat{y}_0$$

$$\lambda, \nu, u \geq 0 \quad (17)$$

Model (17) is equivalent to model (15), so the efficiency scores they produce will be identical.

Theorem 2: For the same data set, model (15) is equivalent to model (18) shown below

$$\min \quad \theta = \frac{1 - \beta \frac{1}{m} \sum_{i=1}^m v_i}{1 + \beta \frac{1}{q} \sum_{r=1}^q u_r}$$

$$\text{s.t.} \quad \hat{X} \lambda + \beta \nu \leq \hat{x}_0$$

$$\hat{Y} \lambda - \beta u \geq \hat{y}_0$$

$$\lambda, \nu, u \geq 0 \quad (18)$$

Proof: Using normalized data, the inputs and outputs of the evaluated DMUs are all equal to unity. We know from the constraint condition of model (18) that

$$0 \leq \beta \leq 1 / \max(v_1, \dots, v_i), \quad i = 1, 2, \dots, m$$

Considering the interval of β , the numerator in model (18) is a monotonic decreasing function, while the denominator is a monotonic increasing function. As a result, within the interval of β , θ is regarded as a monotonic decreasing function. Therefore, model (18) is equivalent to model (15).

Acknowledging that model (18) is a nonlinear programming model, model (15) should be used instead for the measurement of efficiency when the directional distance function is incorporated. On the basis of model (15) we can introduce weights to inputs and outputs according to their relative significance in the efficiency measurement. To be more precise, model (19) is presented

$$\theta = \frac{1 - \beta \frac{1}{m} \sum_{i=1}^m w_i v_i}{1 + \beta \frac{1}{q} \sum_{r=1}^q h_r u_r}$$

$$\begin{aligned}
& \max \beta \\
& \text{s.t. } \hat{X}\lambda + \beta v \leq \hat{x}_0 \\
& \hat{Y}\lambda - \beta u \geq \hat{y}_0 \\
& \lambda, v, u \geq 0 \\
& \sum_{i=1}^m w_i = m, \quad \sum_{r=1}^q w_r = q \quad (19)
\end{aligned}$$

where w stands for the weight assigned to inputs, and h indicates the weight of outputs.

Efficiency measurement can be extended to cases with undesirable outputs. Namely, when undesirable outputs are present, the directional distance function model is defined as follows

$$\begin{aligned}
\theta &= \frac{1 - \frac{1}{m}\beta \sum_{i=1}^m w_i v_i}{1 + \omega \frac{1}{q}\beta \sum_{r=1}^q h_r u_r + \omega' \frac{1}{q'}\beta \sum_{r'=1}^{q'} h_{r'} u_{r'}} \\
& \max \beta \\
& \text{s.t. } X\lambda + \beta v \leq x_k \\
& Y\lambda - \beta u \geq y_k \\
& Y'\lambda + \beta u' \leq y'_k \\
& \sum_{i=1}^m w_i = m, \quad \sum_{r=1}^q w_r = q, \quad \sum_{r'=1}^{q'} w_{r'} = q' \\
& \omega + \omega' = 1 \quad (20)
\end{aligned}$$

where q' denotes the number of undesirable outputs incorporated in the model, h' stands for the weight of undesirable outputs, u' expresses the direction vector of undesirable outputs, and ω and ω' are the weights that determine the mix of desirable and undesirable outputs, respectively, in the measurement of efficiency.

4. Concluding remarks

Although units-invariance is commonly recognized as one of the most fundamental properties of DEA, some DEA models violate this property. The data normalization method we developed in this paper provides a universal solution for this problem. As inputs and outputs are rendered dimensionless, efficiency scores are independent of the measurement units of the inputs and outputs. The proposed data normalization method extends the applicability of the directional distance function model because it eliminates its units-variant problem. However, the virtues of the proposed approach are not limited to the directional distance function, since it can support any future development of DEA that may not respect the units-invariant property.

References

- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30, 1078-1092.
- Chambers, R. G., Chung, Y., & Färe, R. (1996). Benefit and Distance Functions. *Journal of Economic Theory*, 70, 407-419.
- Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory and Applications*, 98, 351-364.
- Charnes, A. (1994). *Data envelopment analysis: theory, methodology, and application*: Kluwer Academic Publishers.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429-444.
- Chung, Y. H., Färe, R., & Grosskopf, S. (1997). Productivity and undesirable outputs: A directional distance function approach. *Journal of Environmental Management*, 51, 229-240.
- Cook, W. D., & Seiford, L. M. (2009). Data envelopment analysis (DEA) - Thirty years on. *European Journal of Operational Research*, 192, 1-17.
- Färe, R., & Knox Lovell, C. A. (1978). Measuring the technical efficiency of production. *Journal of Economic Theory*, 19, 150-162.
- Lovell, C. A. K., & Pastor, J. T. (1995). Units invariant and translation invariant DEA models. *Operations Research Letters*, 18, 147-151.
- Seiford, L. M. (1996). Data envelopment analysis: The evolution of the state of the art (1978–1995) *The Journal of Productivity Analysis*, 6, 99-137.
- Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, 130, 498-509.