Data Envelopment Analysis with Imprecise Data Revisited

Mohammad Izadikhah[a], Dimitris K. Despotis[b]

[a]*Department of Mathematics, College of Science, Arak Branch, Islamic Azad University, P.O. Box: 38135/567, Arak, Iran*

[b]*Department of Informatics, University of Piraeus, Greece*

Article Info

Keywords

Data envelopment analysis,
Interval data,
Interval efficiency.

Article history

Received: 11 March 2020

Accepted: 10 June 2020

Abstract

Wang et al. (2005) proposed a pair of data envelopment analysis (DEA) models to deal with the efficiency assessment of decision-making units (DMU) in the presence of interval input/output data. Their approach was developed with reference to an earlier approach proposed by Despotis and Smirlis (2002) for the same problem. Given that the input/output data are given as interval numbers, the efficiency scores are interval measures as well. In such a setting, both approaches provide lower and upper bounds for the efficiency scores. Wang et al. (2005) claim that the lower and upper bounds calculated in Despotis and Smirlis (2002) are incorrect. Then, they present different models to calculate the true bounds. In this paper, we counter-argue their claim and we show that the Despotis and Smirlis bounds are correct and those provided in Wang et al. are estimated in a manner that they fail to satisfy an obvious property that they should possess. We illustrate our arguments with a counterexample that was originally used in Wang et. al (2005).

1. Introduction

Data envelopment analysis (DEA) (Charnes et al, 1978) is a non-parametric method to measure the efficiency of decision-making units (DMUs) in the presence of multiple inputs and outputs. Standard DEA assumes that all inputs and outputs are exact non-negative ratio estimates. An extension of standard DEA that has gained attention in the literature is the Imprecise DEA (Cooper et al. 1999), where the input/output data are assumed to be mixtures of exact, interval and ordinal data (imprecise data). In the case of the interval data particularly, it is assumed that, due to uncertainty, the true input/output data are only known to lie within bounded intervals. In

*Corresponding author's E-mail: m-izadikhah@iau-arak.ac.ir

such a data setting, the efficiency scores assessed are interval measures as well. Despotis and Smirlis (2002) and then Wang et al. (2005) presented two different pair of models to estimate the upper and the lower bounds of the efficiency scores. Wang et al. (2005) claim that the lower and upper bounds calculated in Despotis and Smirlis (2002) are incorrect. In this paper, we argue and prove that the bounds provided in Wang et al. (2005) are incorrect as they fail to satisfy an obvious property that they should possess.

The note is organized as follows: In Section 2, we present the Despotis and Smirlis (2002) and Wang et al (2005) models for DEA with interval data. In Section 3, we underline the rationale of the two approaches, and we use a simple numerical example to show that the approach of Wang et al. suffers from a serious drawback. Concluding remarks are presented in Section 4.

2. Background

Assume n DMUs, each of which uses m inputs to produce s outputs. We denote by y_{rj} the level of output r ($r = 1, \dots, s$) from DMU j ($j = 1, \dots, n$) and by x_{ij} the level of input i ($i = 1, \dots, m$) used by the DMU j. The standard DEA model for evaluating the efficiency of DMU_p is as follows (input orientation and constant returns-to-scale are assumed):

$$e_p = max \sum_{r=1}^{s} u_r y_{rp}$$

$s.t.$

$$\sum_{i=1}^{m} v_i x_{ip} = 1 \qquad (1)$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \qquad j = 1, \dots, n;$$

$$u_r, v_i \geq 0, \forall r, i$$

Unlike the standard DEA model, it is further assumed that the levels of inputs and outputs are only known to lie within bounded intervals, i.e. $x_{ij} \in [x_{ij}^L, x_{ij}^U]$ and $y_{rj} \in [y_{rj}^L, y_{rj}^U]$, where the upper and lower bounds of the intervals are strictly positive constants. In such a setting, the efficiencies of the DMUs are interval measures as well. Despotis and Smirlis (2002) and Wang et al. (2005), whose models are outlined below, provide in their works DEA models to assess upper and lower bounds for the efficiencies of the DMUs.

The rationale of the interval data in DEA is that, due to uncertainty at the time of assessment, there is lack of knowledge of the exact values of the inputs consumed and the outputs produced but it is possible to provide lower and upper bounds for these values. In that sense, every combination of exact data drawn from within the respective intervals can be assumed as a possible realization of the input/output measures. Let us call such a combination of exact input/output data for any DMU an “occurrence” of that DMU. Obviously, applying standard DEA on a set of occurrences of the DMUs, the exact efficiency scores assessed must lie within the respective efficiency intervals.

2.1 The Despotis and Smirlis (2002) models for DEA with interval data

Despotis and Smirlis (2002) proposed the following pair of DEA models to assess the upper and lower bounds of the evaluated DMU p (the linear forms are given):

$$
\begin{aligned}
E_p^U &= max \sum_{r=1}^{s} u_r y_{rp}^U \\
s.t.& \\
&\sum_{i=1}^{m} v_i x_{ip}^L = 1 \\
&\sum_{r=1}^{s} u_r y_{rp}^U - \sum_{i=1}^{m} v_i x_{ip}^L \leq 0 \\
&\sum_{r=1}^{s} u_r y_{rj}^L - \sum_{i=1}^{m} v_i x_{ij}^U \leq 0, \quad j = 1, \dots, n; j \neq p
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
E_p^L &= max \sum_{r=1}^{s} u_r y_{rp}^L \\
s.t.& \\
&\sum_{i=1}^{m} v_i x_{ip}^U = 1 \\
&\sum_{r=1}^{s} u_r y_{rp}^L - \sum_{i=1}^{m} v_i x_{ip}^U \leq 0 \\
&\sum_{r=1}^{s} u_r y_{rj}^U - \sum_{i=1}^{m} v_i x_{ij}^L \leq 0, \quad j = 1, \dots, n; j \neq p
\end{aligned}
\tag{3}
$$

$u_r, v_i \geq 0, \forall r, i$ $u_r, v_i \geq 0, \forall r, i$

Model (2) provides the upper bound of efficiency E_p^U for the evaluated DMU p, by assessing the efficiency of the best occurrence $\left(x_{ip}^L, y_{rp}^U\right)$ of DMU p against the worst occurrences of all the other DMUs $\left(x_{ij}^U, y_{rj}^L\right), j = 1, \ldots, n; j \neq p$. Model (3), on the other hand, provides the lower bound of efficiency E_p^L for DMU p, by assessing the efficiency of its worst occurrence $\left(x_{ip}^U, y_{rp}^L\right)$ against the best occurrences of all the other DMUs $\left(x_{ij}^L, y_{rj}^U\right), j = 1, \ldots, n; j \neq p$.

We recall here that, applying the standard DEA model (1) on a set of occurrences of the DMUs, the exact efficiency scores assessed must lie within the respective efficiency intervals. The following theorem secures that the bounds estimated by the models (2) and (3) satisfy this property.

Theorem 1: For the efficiency score e_p obtained by assessing any occurrence of DMU p: ($x_{ip} \in \left[x_{ip}^L, x_{ip}^U\right]$, $y_{rp} \in \left[y_{rp}^L, y_{rp}^U\right]$) against any occurrence of the other DMUs: ($x_{ij} \in \left[x_{ij}^L, x_{ij}^U\right]$, $y_{rj} \in \left[y_{rj}^L, y_{rj}^U\right]$) holds that $e_p \in \left[E_p^L, E_p^U\right]$.

Proof.

Firstly, we prove $E_p^L \leq e_p$. Let (u', v') be the optimal solution of model (3) and (u^*, v^*) an optimal solution of model (1). From the first two constraints of model (3) we get $\sum_{r=1}^{s} u'_r y_{rp}^L \leq 1$ and, thus, at optimality we have

$$\sum_{r=1}^{s} u'_r y_{rp}^L \leq 1;$$
$$\sum_{r=1}^{s} u'_r y_{rj}^U - \sum_{i=1}^{m} v'_i x_{ij}^L \leq 0; \quad \forall j, j \neq p$$

According to the following relation

$$\sum_{r=1}^{s} u'_r y_{rj} \leq \sum_{r=1}^{s} u'_r y_{rj}^U \leq \sum_{i=1}^{m} v'_i x_{ij}^L \leq \sum_{i=1}^{m} v'_i x_{rj}$$

We have $\sum_{r=1}^{s} u'_r y_{rj} - \sum_{i=1}^{m} v'_i x_{rj} \le 0,\ \ (j \neq p)$.

Also, from the constraint $\sum_{i=1}^{m} v'_i x_{ip}^U = 1$, we conclude that $\sum_{i=1}^{m} v'_i x_{ip} \le \sum_{i=1}^{m} v'_i x_{ip}^U = 1$. Now, for the constraint $\sum_{i=1}^{m} v_i x_{ip} = 1$ of Model (1), we consider two cases:

Case 1: $\sum_{i=1}^{m} v'_i x_{ip} = 1$.

In this case, $\sum_{r=1}^{s} u'_r y_{rp} \le 1$, and hence[1] we have $\sum_{r=1}^{s} u'_r y_{rp} - \sum_{i=1}^{m} v'_i x_{rp} \le 0$. Therefore, the vector $(\boldsymbol{u}', \boldsymbol{v}')$ is a feasible solution for model (1) and we have

$$E_p^L = \sum_{r=1}^{s} u'_r y_{rp}^L \le \sum_{r=1}^{s} u'_r y_{rp} \le \sum_{r=1}^{s} u^*_r y_{rp} = e_p$$

Case 2: $\sum_{i=1}^{m} v'_i x_{ip} \neq 1$.

Let $\sum_{i=1}^{m} v'_i x_{ip} = q < 1$. We define $\boldsymbol{u}'' = \frac{\boldsymbol{u}'}{q}$ and $\boldsymbol{v}'' = \frac{\boldsymbol{v}'}{q}$. Hence, we have $\sum_{r=1}^{s} u'_r y_{rp} \le \sum_{r=1}^{s} u''_r y_{rp}$, and also

$$\begin{aligned} &\sum_{i=1}^{m} v''_i x_{ip} = 1 \\ &\sum_{r=1}^{s} u''_r y_{rj} - \sum_{i=1}^{m} v''_i x_{ij} \le 0, \qquad (j \neq p) \end{aligned}$$

And similar to the case 1, we have $\sum_{r=1}^{s} u''_r y_{rp} - \sum_{i=1}^{m} v''_i x_{ip} \le 0$. Then, $(\boldsymbol{u}'', \boldsymbol{v}'')$ is a feasible solution for model (1), and so, in this case we also have

$$E_p^L = \sum_{r=1}^{s} u'_r y_{rp}^L \le \sum_{r=1}^{s} u''_r y_{rp} \le \sum_{r=1}^{s} u^*_r y_{rp} = e_p$$

Now, we prove $e_p \le E_p^U$. Let $(\bar{u}, \bar{v})$ be the optimal solution of model (2) and (u^*, v^*) an optimal solution of model (1). In optimality of model (1) we have

[1] Otherwise, if $\sum_{r=1}^{s} u'_r y_{rp} > 1$, we can set $\hat{\boldsymbol{u}} = \boldsymbol{u}' - \boldsymbol{\nabla}$, where $\boldsymbol{\nabla} \in \mathbb{R}^s$ is a vector such that $\sum_{r=1}^{s} \hat{u}_r y_{rp} \le 1$. So, we also have $\sum_{r=1}^{s} \hat{u}_r y_{rp} - \sum_{i=1}^{m} v'_i x_{rp} \le 0$. Hence, the feasibility still remains.

$$\sum_{i=1}^{m} v_i^* x_{ip} = 1;$$
$$\sum_{r=1}^{s} u_r^* y_{rj} - \sum_{i=1}^{m} v_i^* x_{ij} \le 0; \quad \forall j$$

According to the following relation

$$\sum_{r=1}^{s} u_r^* y_{rj}^L \le \sum_{r=1}^{s} u_r^* y_{rj} \le \sum_{i=1}^{m} v_i^* x_{ij} \le \sum_{i=1}^{m} v_i^* x_{ij}^U$$

We have $\sum_{r=1}^{s} u_r^* y_{rj}^L - \sum_{i=1}^{m} v_i^* x_{ij}^U \le 0$ $(j \ne p)$. Now, for the constraint $\sum_{i=1}^{m} v_i x_{ip}^L = 1$ of Model (2), we consider two cases:

Case 1: $\sum_{i=1}^{m} v_i^* x_{ip}^L = 1$.

Then, since we have $\sum_{r=1}^{s} u_r^* y_{rp}^U \le$ [2] $1 = \sum_{i=1}^{m} v_i^* x_{ip}^L$, then clearly we have $\sum_{r=1}^{s} u_r^* y_{rp}^U - \sum_{i=1}^{m} v_i^* x_{ip}^L \le 0$, that shows the vector $(\boldsymbol{u}^*, \boldsymbol{v}^*)$ is a feasible solution for model (2) and we have

$$e_p = \sum_{r=1}^{s} u_r^* y_{rp} \le \sum_{r=1}^{s} u_r^* y_{rp}^U \le \sum_{r=1}^{s} \bar{u}_r y_{rp}^U = E_p^U$$

Where “_” stands for optimality in model (2).

Case 2: $\sum_{i=1}^{m} v_i^* x_{ip}^L < 1$.

Let $\sum_{i=1}^{m} v_i^* x_{ip}^L = h < 1$. We define $\bar{\bar{\boldsymbol{u}}} = \frac{\boldsymbol{u}^*}{h}$ and $\bar{\bar{\boldsymbol{v}}} = \frac{\boldsymbol{v}^*}{h}$. Considering case 1, again we have

$$\sum_{i=1}^{m} \bar{\bar{v}}_i x_{ip}^L = 1,$$
$$\sum_{r=1}^{s} \bar{\bar{u}}_r y_{rp}^U - \sum_{i=1}^{m} \bar{\bar{v}}_i x_{ip}^L \le 0,$$
$$\sum_{r=1}^{s} \bar{\bar{u}}_r y_{rj}^L - \sum_{i=1}^{m} \bar{\bar{v}}_i x_{ij}^U \le 0,$$

That indicate $(\bar{\bar{\boldsymbol{u}}}, \bar{\bar{\boldsymbol{v}}})$ is a feasible solution for model (2). Hence, we have

[2] In the case $\sum_{r=1}^{s} u_r^* y_{rp}^U > 1$, we simply set $\tilde{\boldsymbol{u}} = \boldsymbol{u}^* - \boldsymbol{\Delta}$, where $\boldsymbol{\Delta} \in \mathbb{R}^s$ is a vector such that $\sum_{r=1}^{s} \tilde{u}_r y_{rp}^U \le 1$. So, we also have $\sum_{r=1}^{s} \tilde{u}_r y_{rj}^L - \sum_{i=1}^{m} v_i^* x_{ij}^U \le 0$. Hence, the feasibility still remains.

$$e_p = \sum_{r=1}^{s} u_r^* y_{rp} \leq \sum_{r=1}^{s} u_r^* y_{rp}^U \leq \sum_{r=1}^{s} \bar{\bar{u}}_r y_{rp}^U \leq \sum_{r=1}^{s} \bar{u}_r y_{rp}^U = E_p^U$$

Therefore, for each occurrence of DMU p assessed against any occurrence of the other DMUs holds that $e_p \in [E_p^L, E_p^U]$. This completes the proof. □

2.2 The Wang et al. (2005) models for DEA with interval data

Wang et al. (2005) proposed the following pair of DEA models to assess the upper and lower bounds of the evaluated DMU p:

$$\begin{aligned} \theta_p^U = max \sum_{r=1}^{s} u_r y_{rp}^U & \\ s.t. & \\ \sum_{i=1}^{m} v_i x_{ip}^L = 1 & \\ \sum_{r=1}^{s} u_r y_{rj}^U - \sum_{i=1}^{m} v_i x_{ij}^L \leq 0, & \quad j = 1, \ldots, n \\ u_r, v_i \geq 0, \forall r, i & \end{aligned} \qquad (4)$$

$$\begin{aligned} \theta_p^L = max \sum_{r=1}^{s} u_r y_{rp}^L & \\ s.t. & \\ \sum_{i=1}^{m} v_i x_{ip}^U = 1 & \\ \sum_{r=1}^{s} u_r y_{rp}^L - \sum_{i=1}^{m} v_i x_{ip}^U \leq 0 & \\ \sum_{r=1}^{s} u_r y_{rj}^U - \sum_{i=1}^{m} v_i x_{ij}^L \leq 0, & \quad j = 1, \ldots, n \\ u_r, v_i \geq 0, \forall r, i & \end{aligned} \qquad (5)$$

Model (4) is a standard DEA model applied on input/output bundle represented by the best occurrences of all the DMUs. That is, to obtain the upper bound of efficiency θ_p^U, the best occurrence (x_{ip}^L, y_{rp}^U) of DMU p is assessed against the best occurrences of all the other DMUs $(x_{ij}^L, y_{rj}^U), j = 1, \ldots, n$. Model (5), on the other hand, provides the lower bound of efficiency θ_p^L for DMU p, by assessing the efficiency of its worst occurrence (x_{ip}^U, y_{rp}^L), again against the best occurrences of all the other DMUs, i.e. with respect to the efficiency frontier obtained in model (4). The reader is resorted to Wang et al. (2005), *p.p.* 354, for the interpretation and the argumentation in favor of these efficiency bounds. The structural difference between models (2)-(3) and (4)-(5) is that the latter use a common efficiency frontier to assess the upper and lower bounds for all the DMUs whereas in the former, different frontiers are used, as illustrated in Fig.1. This is the point criticized in Wang et al. and presented as a drawback of Despotis and Smirlis (2002) formulations.

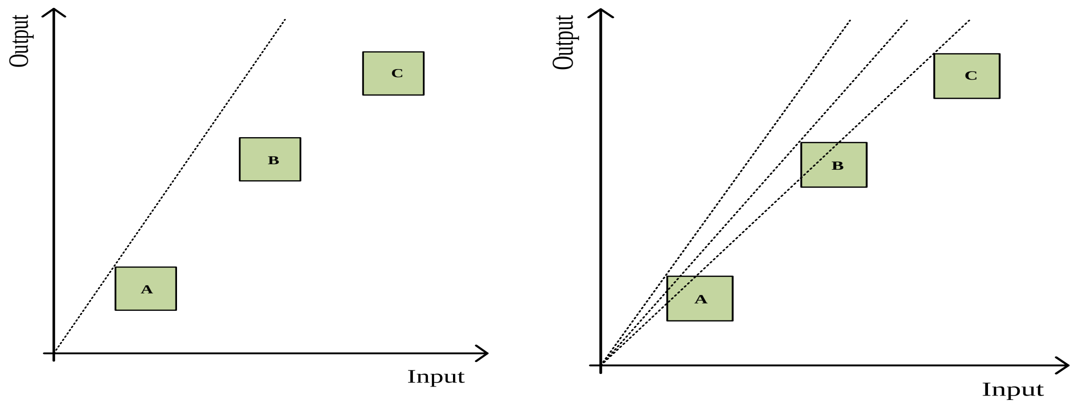
Although the second constraint in (5) is redundant, and therefore correctly is omitted in Wang et al (2005), we keep it to emphasize that the worst occurrence of DMU p, is assessed, among others, against the best occurrence of itself. It is not easily understood how two different realizations of the same DMU can be incorporated simultaneously in the same assessment exercise.

A serious drawback in models (4)-(5) is that for the efficiency interval $[\theta_p^L, \theta_p^U]$ they provide, the theorem 1 does not hold. That is, there might be occurrences of the DMUs for which the exact efficiency score falls outside of the interval $[\theta_p^L, \theta_p^U]$. This is shown with the following counterexample, which is taken from Wang et al. (2005).

Example: Table 1 exhibits three DMUs, A, B and C using one input to produce one output. The input and the output are given as interval numbers. The fourth and the fifth columns present the lower and upper bounds (interval efficiencies) of the three DMUs derived by the models (4)-(5) and (2)-(3) respectively.

Table 3: Interval data and interval efficiencies

DMU	Input $[x_p^L, x_p^U]$	Output $[y_p^L, y_p^U]$	$[\theta_p^L, \theta_p^U]$	$[E_p^L, E_p^U]$
A	[1,2]	[1,2]	[0.25,1.00]	[0.30, 1.00]
B	[3,4]	[4,5]	[0.50, 0.83]	[0.50, 1.00]
C	[5,6]	[6,7]	[0.50, 0.70]	[0.50, 1.00]



a) Wang et al. efficiency frontier b) Despotis and Smirlis efficiency frontiers

Fig. 1: Illustration of interval DMUs (source: Wang et al., 2005)

Consider now an occurrence of the DMUs with exact data drawn from within the interval inputs and outputs as shown in Table 2. The last column of Table 4 shows the efficiency scores derived by applying the standard input-oriented DEA model with constant returns-to-scale (CRS). The efficiency score of DMU C lies outside the interval efficiency [0.50, 0.70], as calculated by models (4)-(5).

Table 2: First occurrence of the DMUs with exact data and efficiency scores

DMU	Input	Output	Efficiency
A	1.00	1.50	1.00
B	3.00	4.50	1.00
C	5.50	6.50	0.79

Moreover, according to Theorem 1, there is not any occurrence of the DMUs that gives to DMU A an efficiency score less than the lower bound 0.30 obtained from model (3). Notice, however, that model (5) set the lower bound for DMU A to 0.25, a value which is unattainable. Thus, an effortless conclusion is that the models (4)-(5) fail to assess the correct bounds. Now, we examine another case that emphasizes the wrong lower bound efficiency score of Wang (2005) model for DMU A (see Table 3).

Table 3: Second occurrence of the DMUs with exact data and efficiency scores

DMU	Input	Output	Efficiency
A	2.00	1.00	0.30
B	3.00	5.00	1.00
C	5.00	7.00	0.84

The case that is reported in Table 3 is the worst case for DMU A and best cases for other DMUs. The result of conventional CCR model shows that its efficiency score is 0.30 and not 0.25. So, there isn't any occurrences related to DMU A with efficiency score lower than 0.30.

3. Conclusion

We presented in this note a theorem and counterexample to prove that, unlike the approach of Despotis and Smirlis (2002), the approach of Wang et al. (2005) to the problem of measuring the performance of DMUs in the presence of interval data fails to estimate the correct efficiency lower and upper bounds of the efficiency scores.

References

[1] Charnes, A., Cooper, W. W., Rhodes, E., Measuring the efficiency of decision making units. European Journal of Operational Research, 1978; 2(6): 429-444.

[2] Cooper, W.W. Park, K.S. & Yu, G. IDEA and AR-IDEA: Models for dealing with imprecise data in DEA, Management Science, 1999; 45: 597–607.

[3] Despotis, D. K., & Smirlis, Y. G., Data envelopment analysis with imprecise data. European Journal of Operational Research, 2002; 140(1): 24-36.

[4] Wang, Y.-M., Greatbanks, R., & Yang, J.-B., Interval efficiency assessment using data envelopment analysis. Fuzzy Sets and Systems, 2005; 153(3): 347-370.