A Novel Data Envelopment Analysis Ranking Based on a Robust Approach

Milan Hladík

Department of Applied Mathematics
Faculty of Mathematics and Physics,
Charles University in Prague, Czech Republic
http://kam.mff.cuni.cz/~hladik/

15th International Conference on Data Envelopment Analysis
DEA 2017, Prague, Czech Republic,
June 26–29, 2017

CCR model for DEA

CCR model for DEA

The classical CCR model (Charnes et al., 1978) for ranking the decision making unit DMU_0 can be formulated as a linear program

$$\max \; y_0^T u \;\; \text{subject to} \;\; x_0^T v \leq 1, \;\; Yu - Xv \leq 0, \;\; u, v \geq 0,$$

where

- $x_0 \in \mathbb{R}^{n_1}$ is the input nonnegative vector for DMU_0,
- $y_0 \in \mathbb{R}^{n_2}$ is the output nonnegative vector for DMU_0,
- $X \in \mathbb{R}^{m \times n_1}$ is the input nonnegative matrix for the other DMU's, in particular, the ith row of X is the input vector for the ith DMU,
- $Y \in \mathbb{R}^{m \times n_2}$ is the output nonnegative matrix for the other DMU's, in particular, the ith row of Y is the output vector for the ith DMU,
- u and v are vectors of variables representing output and input weights, respectively.

Our goal

Discussion

Many other models exist.

Our goal

Introduce a new efficiency ranking based on a robustness point of view.

Why?

Besides the positive properties of the classical model:

- quantify stability/robustness
- measure distance to (in)efficiency
- handle precise as well as imprecise data

Other approaches

- robust models for imprecise data

Novel ranking – idea

The idea of novel robust ranking

- determine the largest allowable variations of the input and output data such that DMU_0 remains efficient (for efficient DMU's), or
- the smallest possible variation of the input and output data such that DMU_0 becomes efficient (for inefficient DMU's)
- the corresponding coefficient of variations gives us a new ranking

Novel ranking – formalization

δ-neighborhood

Define δ-neighborhood of the data as

$$\begin{aligned}\mathcal{O}_\delta(x_0, y_0, X, Y) = \{(x_0', y_0', X', Y') \colon & |x_{ij}' - x_{ij}| \le \delta x_{ij}, \\ & |y_{ik}' - y_{ik}| \le \delta y_{ik},\ \forall i, j, k\}.\end{aligned}$$

New ranking

If DMU_0 is efficient, then its ranking is defined as $r = 1 + \delta^*$, where

$$\begin{aligned}\delta^* = \max\{\delta \colon & \text{DMU}_0 \text{ is efficient for all} \\ & (x_0', y_0', X', Y') \in \mathcal{O}_\delta(x_0, y_0, X, Y)\}.\end{aligned}$$

If DMU_0 is inefficient, then its ranking is defined as $r = 1 + \delta^*$, where

$$\begin{aligned}\delta^* = -\min\{\delta \colon & \text{DMU}_0 \text{ is efficient for some} \\ & (x_0', y_0', X', Y') \in \mathcal{O}_\delta(x_0, y_0, X, Y)\}.\end{aligned}$$

Notice that the maximum (minimum) value needn't be attained.

Novel ranking – computation

Theorem

We have

$$\delta^* = \max\ \delta \quad \textit{subject to} \quad (1-\delta) y_0^T u \geq 1,\ (1+\delta) x_0^T v \leq 1,$$
$$(1+\delta) Yu - (1-\delta) Xv \leq 0,\ u, v \geq 0.$$

Properties

- it is a nonlinear programming problem in variables δ, u, v
- still efficiently computable
- in a class of generalized linear fractional programming problems, which have the form of

$$\min\ \lambda \quad \text{subject to} \quad Ax \leq \lambda Bx,\ Cx \leq c,\ x \geq 0,$$

where $Bx \geq 0$ holds for all x satisfying $Cx \leq c$, $x \geq 0$. Equivalently

$$\min\ \left(\max_i \frac{(Ax)_i}{(Bx)_i}\right) \quad \text{subject to} \quad Cx \leq c,\ x \geq 0.$$

Novel ranking – computation

The resulting linearized model

The ranking is $r = 1 + \delta^*$, where δ^* is computed by the linear program

$$\delta^* = \max\ \delta \ \text{ subject to } \ y_0^T \tilde{u} \geq 1 + \delta,\ x_0^T \tilde{v} \leq 1 - \delta,$$
$$Y\tilde{u} - X\tilde{v} \leq 0,\ \tilde{u}, \tilde{v} \geq 0.$$

Properties

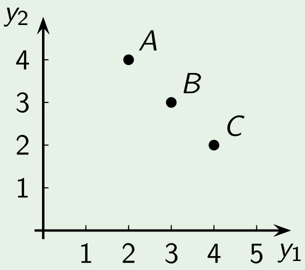
- invariant to scaling the units of input and output data
- $r \in [0, 2]$
- $r \geq 1$ if and only if DMU_0 is efficient
- $r < 1$ if and only if DMU_0 is inefficient
- does not change (in)efficiency of the classical ranking,
  even the order is not changed (for both linear and nonlinear models)

Novel ranking – properties

New ranking as a robustness measure

- we can use r as the measure of inefficiency and efficiency
- if $r = 1 + \delta^* \geq 1$, then DMU_0 is efficient for any variation of the data up to $100\delta^*\%$ of their nominal values; moreover, all data coefficients may vary simultaneously and independently to each other.

Example



Outputs of DMU's A, B and C are

$$A : (2, 4), \quad B : (3, 3), \quad C : (4, 2).$$

- the classical ranks: 1, 1, 1
- a slight change of the outputs of B makes it inefficient
- our ranks are 1.1429, 1, 1.1429
- B is borderline; A, C are stable (under 14.29% perturbation)

Novel ranking – properties

New ranking as a global ranking

- the novel ranking is naturally normalized
- use for comparing DMU's from different, even unrelated models

Example

Suppose that we have a ranking of banks like

$$1.0062,\ 0.986,\ 1.0397,\ 1.024,\ 0.97263,\ 1.0009,\ 1.0438,\ 0.96441,$$

and suppose that we have a ranking of hospitals like

$$1.21,\ 0.65338,\ 1.3254,\ 0.6799,\ 1.0382,\ 0.60379,\ 0.89957,\ 1.2454.$$

- In the first case, all the banks have very similar ranking, no substantial difference in their performance
- performance of particular hospitals differs a lot
- there are some considerably efficient and some highly inefficient
- this shows the universal feature of our approach to ranking

Handling interval data

- suppose we are given interval data $[\underline{x}_0, \overline{x}_0]$, $[\underline{X}, \overline{X}]$, $[\underline{Y}, \overline{Y}]$ and $[\underline{x}_0, \overline{x}_0]$
- the best case rank $\overline{r}$ happens in the setting

$$x_0 := \underline{x}_0, \; y_0 := \overline{y}_0, \; X := \overline{X} \text{ and } Y := \underline{Y}$$

- the worst case rank $\underline{r}$ happens for

$$x_0 := \overline{x}_0, \; y_0 := \underline{y}_0, \; X := \underline{X} \text{ and } Y := \overline{Y}$$

- by solving two real-valued problems we get the efficiency range $[\underline{r}, \overline{r}]$
- if $\underline{r} \geq 1$, then DMU_0 is always efficient
- if $\overline{r} < 1$, then DMU_0 is never efficient

Novel ranking – additional DMU / input or output

Additional DMU

- additional DMU $=$ a novel constraint in the optimization model
- the ranking cannot increase

Theorem

Suppose DMU_0 is efficient. The set of all values $x_a \in \mathbb{R}^{n_1}$ and $y_a \in \mathbb{R}^{n_1}$ causing DMU_0 to be inefficient forms an interior of a convex polyhedron.

Additional inputs or outputs

- the ranking cannot decrease

Theorem

Suppose DMU_0 is inefficient. The set of all values $(y_b, y_c) \in \mathbb{R}^{1+m}$ causing DMU_0 to be efficient forms an interior of a convex polyhedron.

Novel ranking – Example 1

Example (Cooper et al., 2007)

Ranking hospitals:

	INPUT		OUTPUT				
DMU	doctors	nurses	outpatients	inpatients	classical eff.	new eff. (lin.)	new eff.
A	20	151	100	90	1	1.1696	1.1708
B	19	131	150	50	1	1.0843	1.0845
C	25	160	160	55	0.8827	0.9377	0.9376
D	27	168	180	72	1	1.0079	1.0079
E	22	158	94	66	0.7635	0.8659	0.8653
F	55	255	230	90	0.8348	0.9100	0.9097
G	33	235	220	88	0.9020	0.9485	0.9484
H	31	206	152	80	0.7963	0.8866	0.8863
I	30	244	190	100	0.9604	0.9798	0.9798
J	50	268	250	100	0.8707	0.9309	0.9307
K	53	306	260	147	0.9551	0.9770	0.9770
L	38	284	250	120	0.9582	0.9787	0.9787

Novel ranking – Example 2

Example (Entani et al., 2002, He et al., 2016, . . .)

DMU	X_1	Y_1	Y_2	efficiency from Refs	novel efficiency
A	1	[0.8, 1.2]	[7.50, 8.50]	[1, 1]	[1.0169, 1.1148]
B	1	[1.8, 2.2]	[2.50, 3.50]	[0.4222, 0.6227]	[0.5937, 0.7675]
C	1	[1.6, 2.4]	[5.75, 6.25]	[0.7297, 0.9167]	[0.8437, 0.9566]
D	1	[2.5, 3.5]	[2.75, 3.25]	[0.5247, 0.7809]	[0.6882, 0.8770]
E	1	[2.8, 3.2]	[6.75, 7.25]	[0.9646, 1]	[0.9819, 1.1292]
F	1	[3.8, 4.2]	[1.83, 2.17]	[0.6131, 0.7806]	[0.7601, 0.8768]
G	1	[3.4, 4.6]	[4.50, 5.50]	[0.7940, 1]	[0.8852, 1.0643]
H	1	[4.7, 5.3]	[1.50, 2.50]	[0.6984, 0.9635]	[0.8224, 0.9814]
I	1	[5.6, 6.4]	[1.67, 2.33]	[0.8229, 1]	[0.9028, 1.0482]
J	1	[6.7, 7.3]	[0.75, 1.25]	[1, 1]	[1.0229, 1.1318]

- A and J are efficient for each realization
- B, C, D, F, and H are inefficient for each realization
  B or F are far to efficiency while H is possibly closer
- E is either efficient or very close to efficiency for each realization.

Conclusion

Summary

New DEA ranking based on robustness of DMU's of their (in)efficiency with many attractive properties:

- efficiently computable by linear programming
- invariant with respect to scaling
- it gives a measure of efficiency as a distance to inefficiency and vice versa
- suitable as a universal ranking technique of DMU's of different models.
- suitable for further generalization – models with interval data etc.