GRIPS Discussion Paper 13-10

How to deal with S-shaped curve in DEA

Kaoru Tone
Miki Tsutsui

June 2013

GRIPS
NATIONAL GRADUATE INSTITUTE FOR POLICY STUDIES

National Graduate Institute for Policy Studies
7-22-1 Roppongi, Minato-ku,
Tokyo, Japan 106-8677

How to deal with S-shaped curve in DEA

Kaoru Tone[a], Miki Tsutsui[b]

[a]National Graduate Institute for Policy Studies, Tokyo, Japan
[b]Central Research Institute of Electric Power Industry, Tokyo, Japan
tong@grips.ac.jp, miki@criepi.denken.or.jp

Abstract

In DEA we are often puzzled by the big difference in CRS and VRS scores, and by the convex production possibility set syndrome in spite of the S-shaped curve often observed in many real data. In this paper we perform a challenge to these subjects.

Keywords: Data envelopment analysis, S-shaped curve, CRS, VRS, scale elasticity, SAS

1. Motivation

In DEA (Data Envelopment Analysis), we are often puzzled by the big difference between the constant returns-to-scale score (CRS) and the variable returns-to-scale score (VRS). Several authors (Avkiran (2001), Avkiran et al. (2008), Bogetoft and Otto (2010) among others) proposed solutions for this problem. In this paper we propose a different approach and results. Another problem is the conventional convex production possibility set assumption which is closely related to the first problem. In this paper, we discuss these two basic subjects of DEA.

Several researchers have discussed non-convex production possibility set issues, see Dekker and Post (2001), Kousmanen (2001), Podinovski (2004), Olsen and Petersen (2013), among others. However, we believe there is room for further research on this subject.

Another objective of this paper is the measurement of scale elasticity of production. Most of researches on this subject are based on the convex production possibility set assumption. We propose a new scheme for evaluation of scale elasticity within the cluster each DMU belongs to.

This paper unfolds as follows. In Section 2, we describe a decomposition of the CRS slacks after introducing basic notations, and define the scale-independent data set. In Section 3, we introduce clusters and define the scale&cluster-adjusted score (SAS). In Section 4 we explain our scheme using a tiny example. Two illustrative examples are presented in Section 5. In Section 6, we define the scale elasticity based on the scale-dependent data set. An empirical study on Japanese universities follows in Section 7. Extensions to the radial DEA models are presented in Section 8. The last section concludes this paper.

2. Global issue

In this section we introduce notation and basic tools, and discuss a decomposition of slacks.

2.1. Notation and basic tools

Let the input and output data matrices be respectively

$$\begin{aligned}&\mathbf{X} \in R_{+}^{m\times n} (= (x_{ij})\,(i=1,\dots,m;\ j=1,\dots,n)) \text{ and}\\ &\mathbf{Y} \in R_{+}^{s\times n} (= (y_{rj})\,(r=1,\dots,s;\ j=1,\dots,n)),\end{aligned} \tag{1}$$

where m, s and n are the number of inputs, outputs and decision making units (DMUs).

Then, the production possibility set for the constant returns-to-scale (CRS) and variable returns-to-scale (VRS) models are defined respectively by

$$P_{CRS} = \{(\mathbf{x},\mathbf{y}) \mid \mathbf{x} \geq \mathbf{X}\boldsymbol{\lambda}, \mathbf{y} \leq \mathbf{Y}\boldsymbol{\lambda}, \boldsymbol{\lambda} \geq \mathbf{0}\}, \tag{2}$$

$$P_{VRS} = \{(\mathbf{x},\mathbf{y}) \mid \mathbf{x} \geq \mathbf{X}\boldsymbol{\lambda}, \mathbf{y} \leq \mathbf{Y}\boldsymbol{\lambda}, \mathbf{e}\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \mathbf{0}\}, \tag{3}$$

where $\mathbf{x} \in R_{+}^{m}$, $\mathbf{y} \in R_{+}^{s}$ and $\boldsymbol{\lambda}(\geq \mathbf{0}) \in R^{n}$ are input, output, and intensity vectors, and $\mathbf{e} \in R^{n}$ is the row vector with all elements equal to 1.

Throughout this section, we utilize the input-oriented slacks-based measure (SBM) (Tone (2001)) for the efficiency evaluation of each DMU $(x_o, y_o)\,(o = 1,\dots,n)$ regarding the CRS and VRS models as follows:

$$\begin{aligned}&[\text{CRS}]\ \theta_o^{CRS} = \min 1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^-}{x_{io}}\\ &\quad\text{subject to}\\ &\quad\mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^- = \mathbf{x}_o\\ &\quad\mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^+ = \mathbf{y}_o\\ &\quad\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}.\end{aligned} \tag{4}$$

$$\begin{aligned}&[\text{VRS}]\ \theta_o^{VRS} = \min 1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^-}{x_{io}}\\ &\quad\text{subject to}\\ &\quad\mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^- = \mathbf{x}_o\\ &\quad\mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^+ = \mathbf{y}_o\\ &\quad\mathbf{e}\boldsymbol{\lambda} = 1\\ &\quad\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0},\end{aligned} \tag{5}$$

where $\boldsymbol{\lambda} \in R^n$ is the intensity vector and $\mathbf{s}^-$, $\mathbf{s}^+$ are respectively input- and output-slacks.

Although we present our model in the input-oriented SBM model, we can develop the model to the output-oriented and non-oriented SBM models as well as to the radial models (Section 8).

We define the scale-efficiency (σ_o) of DMU$_o$ by

$$\sigma_o = \frac{\theta_o^{CRS}}{\theta_o^{VRS}}. \tag{6}$$

We denote optimal slacks of the CRS model by

$$(\mathbf{s}_o^{-*}, \mathbf{s}_o^{+*}). \tag{7}$$

Although we utilize the scale-efficiency CRS/VRS as an index of scale merits and demerits, we can make use of other indexes appropriate for discriminating handicaps due to scale. However, the index must be normalized between 0 and 1, and the larger indicates the better scale condition.

2.2. Decomposition of CRS slacks

We decompose CRS slacks into scale-independent and –dependent parts as follows:

$$\begin{aligned} \mathbf{s}_o^{-*} &= \sigma_o \mathbf{s}_o^{-*} + (1-\sigma_o)\mathbf{s}_o^{-*} \\ \mathbf{s}_o^{+*} &= \sigma_o \mathbf{s}_o^{+*} + (1-\sigma_o)\mathbf{s}_o^{+*} \end{aligned} \tag{8}$$

If DMUo satisfies $\sigma_o = 1$ (so called in *the most productive scale size*), its slacks are all attributed to the scale-independent slacks. However, if $\sigma_o < 1$, its slacks are decomposed into the scale-independent part $(\sigma_o \mathbf{s}_o^{-*}, \sigma_o \mathbf{s}_o^{+*})$ and the scale-dependent part $((1-\sigma_o)\mathbf{s}_o^{-*}, (1-\sigma_o)\mathbf{s}_o^{+*})$.

$$\text{Scale-independent slacks} = (\sigma_o \mathbf{s}_o^{-*}, \sigma_o \mathbf{s}_o^{+*}) \tag{9}$$

$$\text{Scale-dependent slacks} = ((1-\sigma_o)\mathbf{s}_o^{-*}, (1-\sigma_o)\mathbf{s}_o^{+*}). \tag{10}$$

2.3. Scale-independent data set

We define the scale-independent data $(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$ $(o = 1, \ldots, n)$ by deleting and adding the scale-depending slacks as:

$$\begin{array}{ll} \text{Scale-independent Input} & \bar{\mathbf{x}}_o = \mathbf{x}_o - (1-\sigma_o)\mathbf{s}_o^{-*} \\ \text{Scale-independent Output} & \bar{\mathbf{y}}_o = \mathbf{y}_o + (1-\sigma_o)\mathbf{s}_o^{+*} \end{array} \tag{11}$$

See Figure 1 for an illustration.

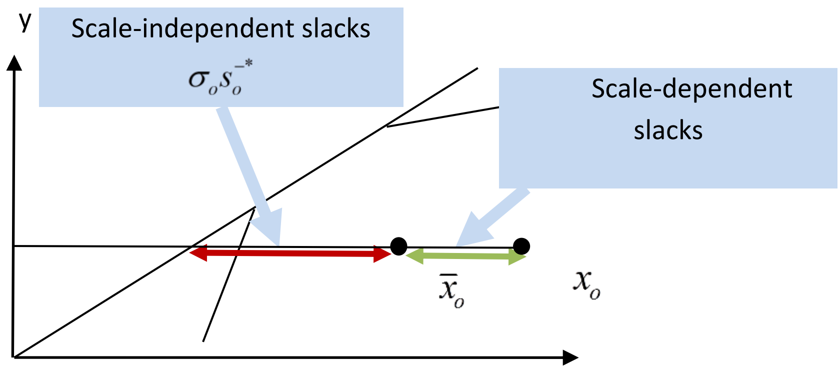

Figure 1: Scale-independent input

3. In-cluster issue: Scale&Cluster-adjusted DEA score (SAS)

In this section we introduce the cluster of DMUs and define the scale&cluster-adjusted score (SAS).

3.1. Cluster

We classify DMUs into several clusters depending on their characteristics. They can be supplied exogenously (see Section 6 for an example), or determined posteriori depending on the degree of scale-efficiency. A sample of the latter case may go as follows. We already know returns-to-scale (RTS) characteristics of each DMU, i.e. IRS , CRS or DRS, from the VRS solution. We first classify CRS DMUs as Cluster C. Then we classify IRS DMUs depending on the degree of scale-efficiency σ. For example, for IRS DMUs with $1 > \sigma \geq 0.8$ we classify them as I1, with $0.8 > \sigma \geq 0.6$ as I2, and so on. For DRS DMUs with $1 > \sigma \geq 0.8$ we classify them as D1, with $0.8 > \sigma \geq 0.6$ as D2, and so on. We must decide the number of clusters and bandwidth considering the number of DMUs.

We denote the name of cluster DMU$_j$ by Cluster(j) $(j = 1,\ldots,n)$.

3.2. Solving the CRS model in the same cluster

We solve the CRS model for each DMU $(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$ $(o = 1,\ldots,n)$ referring to the $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ in the same Cluster (o) which can be formulated as follows:

$$
\begin{aligned}
&\min 1-\frac{1}{m}\sum_{i=1}^{m}\frac{s_i^{cl-}}{\bar{x}_{io}}\\
&\quad\text{subject to}\\
&\quad\bar{\mathbf{X}}\boldsymbol{\mu}+\mathbf{s}^{cl-}=\bar{\mathbf{x}}_o\\
&\quad\bar{\mathbf{Y}}\boldsymbol{\mu}-\mathbf{s}^{cl+}=\bar{\mathbf{y}}_o\\
&\quad\mu_j=0\quad(\forall j:\text{Cluster}(j)\neq\text{Cluster}(o))\\
&\quad\boldsymbol{\mu}\geq\mathbf{0},\mathbf{s}^{cl-}\geq\mathbf{0},\mathbf{s}^{cl+}\geq\mathbf{0}.
\end{aligned}
\tag{12}
$$

We denote an optimal in-cluster slacks by $(\mathbf{s}_o^{cl-*},\mathbf{s}_o^{cl+*})$. By adding the scale-dependent slacks and in-cluster slacks, we define the total slacks as

$$
\begin{aligned}
&\text{Total input slacks}\quad \bar{\mathbf{s}}_o^{-}=(1-\sigma_o)\mathbf{s}_o^{-*}+\mathbf{s}_o^{cl-*}\\
&\text{Total output slacks}\quad \bar{\mathbf{s}}_o^{+}=(1-\sigma_o)\mathbf{s}_o^{+*}+\mathbf{s}_o^{cl+*}
\end{aligned}
\tag{13}
$$

Scale&cluster-adjusted data (projection) $(\bar{\bar{\mathbf{x}}}_o,\bar{\bar{\mathbf{y}}}_o)$ is defined by:

$$
\begin{aligned}
&\text{Scale\&cluster-adjusted input (Projected Input)}\\
&\qquad\bar{\bar{\mathbf{x}}}_o=\mathbf{x}_o-\bar{\mathbf{s}}_o^{-}=\mathbf{x}_o-(1-\sigma_o)\mathbf{s}_o^{-*}-\mathbf{s}_o^{cl-*}\\
&\text{Scale\&cluster-adjusted output (Projected Output)}\\
&\qquad\bar{\bar{\mathbf{y}}}_o=\mathbf{y}_o+\bar{\mathbf{s}}_o^{+}=\mathbf{y}_o+(1-\sigma_o)\mathbf{s}_o^{+*}+\mathbf{s}_o^{cl+*}
\end{aligned}
\tag{14}
$$

See Figure 2 for an illustration.

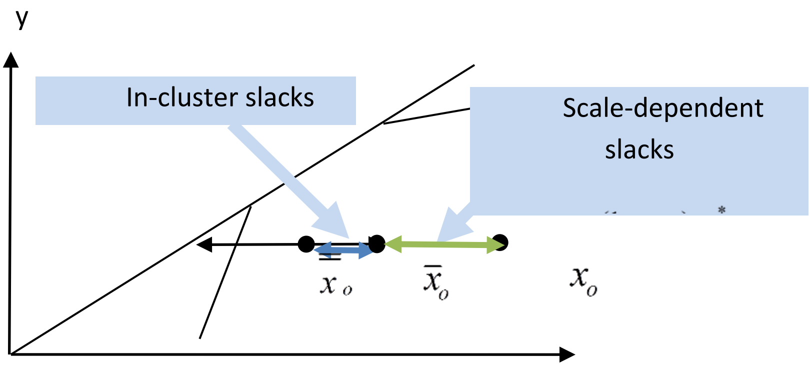

Figure 2: Scale&cluster-adjusted input

Up to this point, we deleted scale demerits and in-cluster slacks from the data set. Thus, we have obtained a scale free and in-cluster slacks free (projected) data set $(\bar{\bar{\mathbf{X}}}, \bar{\bar{\mathbf{Y}}})$.

3.3. Scale&Cluster-adjusted score (SAS)

In the input-oriented case, the scale&cluster-adjusted score (SAS) is defined by

$$\text{Scale\&cluster-adjusted score (SAS)} \quad \theta_o^{SAS} = 1 - \frac{1}{m}\sum_{i=1}^{m} \frac{\bar{s}_{io}^{-}}{x_{io}} = 1 - \frac{1}{m}\sum_{i=1}^{m} \frac{s_{io}^{cl-*} + s_{io}^{-*}}{x_{io}} . \quad (15)$$

The reason why we utilize the above scheme is as follows. First, we wish to eliminate scale demerits from the CRS slacks. For this purpose, we decompose the CRS slacks into scale-dependent and –independent parts, in the recognition of scale demerits as represented by 1-σ_o. If σ_o =1, the DMU has no scale demerits and its slacks are attributed to itself. If σ_o =0.25, then 75% of the slacks are attributed to its scale demerits. After deleting the scale-dependent slacks, we evaluate the DMU within the cluster it belongs to and find in-cluster slacks. If the DMU is efficient among its cluster, its in-cluster slacks are zero, while, if inefficient, the DMU has in-cluster slacks against the efficient DMU. Lastly, we add the in-cluster and scale-dependent slacks to obtain the total slacks. Using the total slacks, we define the scale&cluster-adjusted score (SAS).

[Proposition 1] The scale&cluster-adjusted score (SAS) is not less than the CRS score.

$$\theta_o^{SAS} \geq \theta_o^{CRS} . \quad (16)$$

[Proposition 2] If $\theta_o^{CRS} = 1$ then it holds $\theta_o^{SAS} = \theta_o^{CRS}$, but not vice versa.

[Proposition 3] The scale&cluster-adjusted score (SAS) is decreasing in the increase of input and in the decrease of output so long as the both DMUs remain in the same cluster.

[Proposition 4] The projected DMU $(\bar{\bar{\mathbf{x}}}_o, \bar{\bar{\mathbf{y}}}_o)$ is efficient under the SAS model among the DMUs in the cluster it belongs to. It is also CRS and VRS efficient among the DMUs in its cluster.

All proofs are in Appendix A.

4. How does it work

We demonstrate the above procedure using a tiny example.

Table 1 exhibits 5 DMUs with a single input *x* and a single output *y*. Figure 3 display them where the CRS efficient frontier is the line OA while the VRS efficient lines are AB and BC. We assume DMUs B and D belong to the same cluster b while others belong to themselves.

Table 1: Five DMUs

DMU	(I)x	(O)y	Cluster
A	9	9	a
B	6	4	b
C	5	1	c
D	9	4	b
E	8	5	e

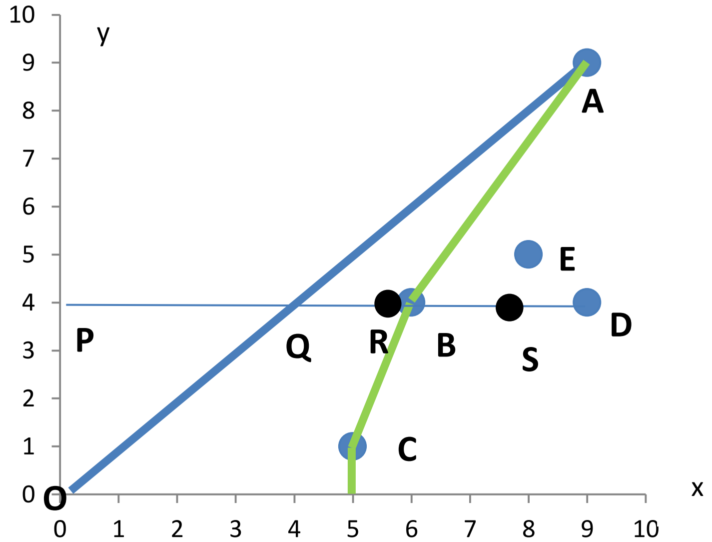

Figure 3: DMUs

For DMU B, we have

$$s_B^- = \text{QB} = 2, \sigma_B = \text{PQ/PB} = 0.6667$$
$$\text{Scale-dependent slack} = \text{RB} = (1-\sigma_B)s_B^- = 0.6667$$
$$\text{In-cluster slack} = 0$$
$$\text{Total slack} = 0.6667.$$

Hence

$$\theta_B^{SAS} = 1 - \frac{\text{RB}}{\text{PB}} = 1 - \frac{0.6667}{6} = 0.8889$$
$$\text{Scale\&cluster-adjusted input } \overset{=}{x}_B = x_B - \text{Total slack} = 5.3333.$$

For DMU D, we have

$$s_D^- = \text{QD} = 5, \sigma_D = \text{PQ/PB} = 0.6667$$
$$\text{Scale-dependent slack} = \text{SD} = (1-\sigma_D)s_D^- = 1.6667$$
$$\text{In-cluster slack} = \text{RS} = 2$$
$$\text{Total slack} = \text{RD} = \text{RS+SD} = 3.6667.$$

In-Cluster slack occurs against DMU B, because B and D belong to the same cluster b. Hence

$$\theta_D^{SAS} = 1 - \frac{\text{RD}}{\text{PD}} = 1 - \frac{3.666}{9} = 0.5926$$
$$\text{Scale\&cluster-adjusted input } \overset{=}{x}_D = x_D - \text{Total slack} = 5.3333.$$

The situation of DMU E differs from other DMUs. This DMU belong to the cluster consisting of itself and is inefficient regarding to both CRS and VRS models. See Figure 4.

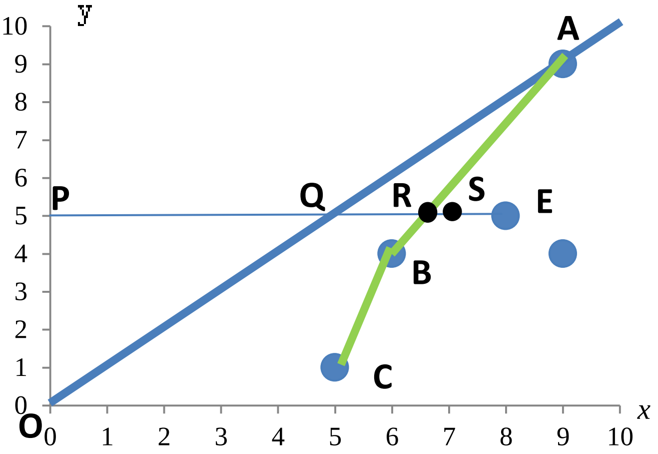

Figure 4: DMU E

DMU E has the following elements:

$$\theta_E^{CRS} = 0.625 : \theta_E^{VRS} = 0.825$$

$$\sigma_E = \frac{\text{PQ}}{\text{PR}} = \frac{5}{6.6} = 0.7576$$

$$s_E^- = \text{QE} = 3$$

$$\text{Scale-dependent slack} = \text{SE} = (1-\sigma_E)s_E^- = 0.7272$$

$$\text{In-cluster slack} = 0$$

$$\text{Total slack } = \text{SE} = 0.7272$$

$$\text{Scale\&cluster-adjusted score } \theta_E^{SAS} = \frac{\text{PS}}{\text{PE}} = 1 - \frac{\text{SE}}{\text{PE}} = 1 - \frac{\text{Total slack}}{x_E} = 0.9091$$

$$\text{Scale\&cluster-adjusted input } \bar{\bar{x}}_E = x_e - \text{Total slack} = 7.2728.$$

DMU E has no In-cluster slack, because it is isolated in cluster. Its Scale&cluster-adjusted score SAS is larger than the VRS score. Table 2 exhibits results of computation and Figure 5 displays Scale&cluster-adjusted projections. Frontiers are non-convex. The non-convexity is caused by the recognition of scale demerits and clusters.

Even when σ_o =1 for all DMUs, clustering may bring non-convex frontiers.

Table 2: Comparisons of three scores with projected input and output

DMU	CRS-I	VRS-I	SAS-I	SAS Projection Input	SAS Projection Output
A	1	1	1	9	9
B	0.6667	1	0.8889	5.3333	4
C	0.2	1	0.36	1.8	1
D	0.4444	0.6667	0.5926	5.3333	4
E	0.625	0.825	0.9091	7.2727	5

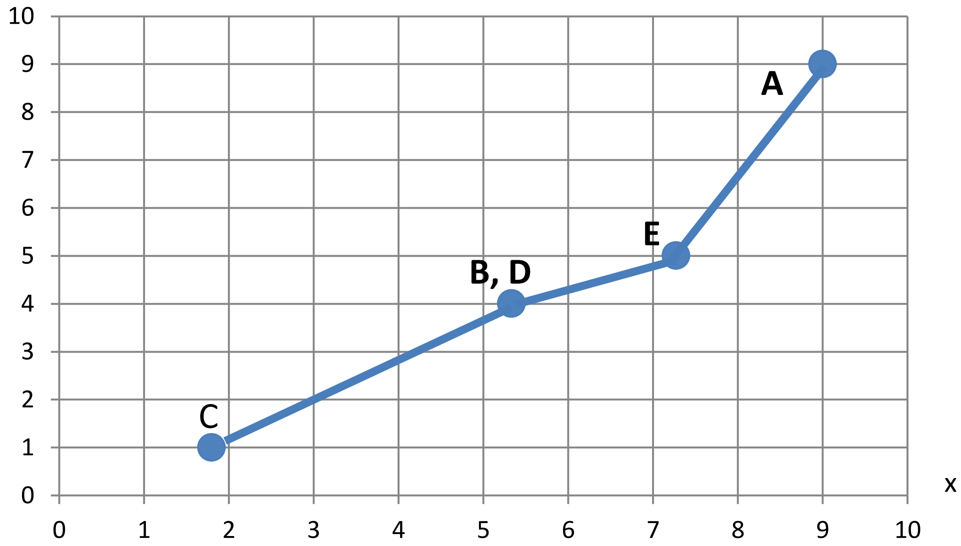

Figure 5: Projected x and y (frontiers)

5. Illustrative examples

In this section we present two artificial examples with a single input and a single output. The first one is totally non-convex, and the second one is a mixture of non-convex and convex frontiers. We demonstrate the above procedures using them.

5.1. Example 1

Table 3 shows 19 DMUs with input *x* and output *y*, while Figure 6 exhibits them graphically. We assume that DMUs A, B and C belong to Cluster a, and DMUs K and L to Cluster k, while other DMUs belong to themselves.

Table 3: Example 1

DMU	(I)x	(O)y	Cluster	DMU	(I)x	(O)y	Cluster
A	2	0.5	a	K	5	5	k
B	3	0.5	a	L	6	5	k
C	3.5	0.6	a	M	7	5.2	m
D	4	1	d	N	7.5	5.3	n
E	4.25	1.5	e	O	8	5.5	o
F	4.5	2	f	P	8.5	5.8	p
G	4.6	2.5	g	Q	9	6.2	q
H	4.7	3	h	R	9.5	6.7	r
I	4.8	3.5	i	S	10	7.3	s
J	4.9	4	j				

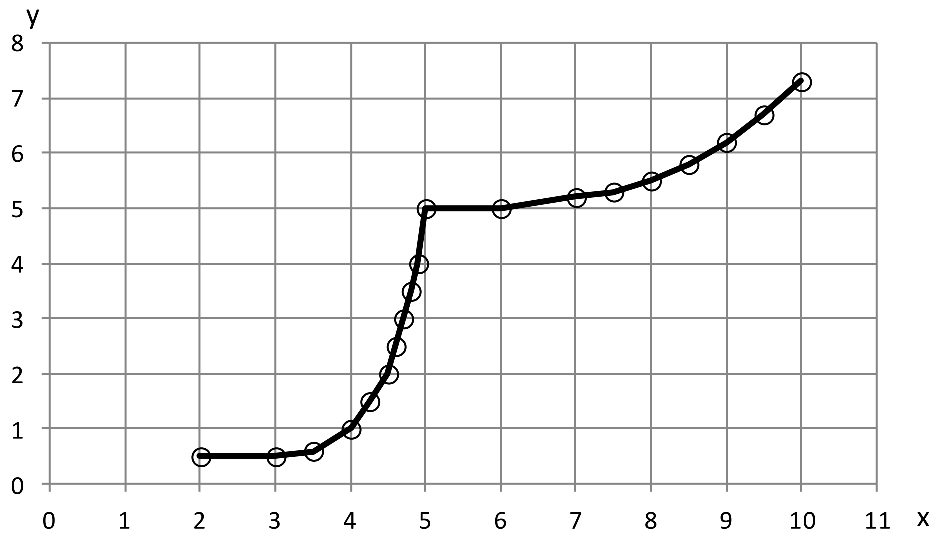

Figure 6: Data plot of Example 1

First, we solved the input-oriented CRS and VRS models, and obtained the scale-efficiency and CRS slacks which were decomposed into the scale-independent and –dependent parts. Table 4 exhibits them. Since the output *y* has no slacks in this example, we do not display them.

Table 4: CRS, VRS, Scale-efficiency and Slacks

				CRS Slacks	Scale-Independent Slacks	Scale-Dependent Slacks
DMU	CRS-I	VRS-I	Scale-Eff.σ	s^-	σs^-	$(1-\sigma)s^-$
A	0.25	1	0.25	1.5	0.375	1.125
B	0.1667	0.6667	0.25	2.5	0.625	1.875
C	0.1714	0.5905	0.2903	2.9	0.8419	2.0581
D	0.25	0.5833	0.4286	3	1.2857	1.7143
E	0.3529	0.6275	0.5625	2.75	1.5469	1.2031
F	0.4444	0.6667	0.6667	2.5	1.6667	0.8333
G	0.5435	0.7246	0.75	2.1	1.575	0.525
H	0.6383	0.7801	0.8182	1.7	1.3909	0.3091
I	0.7292	0.8333	0.875	1.3	1.1375	0.1625
J	0.8163	0.8844	0.9231	0.9	0.8308	0.0692
K	1	1	1	0	0	0
L	0.8333	0.8333	1	1	1	0
M	0.7429	0.7764	0.9568	1.8	1.7222	0.0778
N	0.7067	0.7536	0.9377	2.2	2.0629	0.1371
O	0.6875	0.7609	0.9036	2.5	2.2589	0.2411
P	0.6824	0.7928	0.8606	2.7	2.3237	0.3763
Q	0.6889	0.8454	0.8149	2.8	2.2816	0.5184
R	0.7053	0.9153	0.7705	2.8	2.1574	0.6426
S	0.73	1	0.73	2.7	1.971	0.729

Second, we deleted the scale-dependent slacks from the data and obtained the data set $(\overline{\mathbf{X}},\overline{\mathbf{Y}})$. We solved the CRS model within the same cluster and found the in-cluster slacks. By adding the scale-dependent slacks and in-cluster slacks we obtained the total slacks.

Table 5 records them.

Table 5: $(\overline{\mathbf{X}},\overline{\mathbf{Y}})$, In-cluster slacks and Total slacks

DMU	Cluster	$\bar{x}$	$\bar{y}$	In-cluster slacks	Scale-dependent slacks	Total slacks
A	a	0.875	0.5	0	1.125	1.125
B	a	1.125	0.5	0.25	1.875	2.125
C	a	1.4419	0.6	0.3919	2.0581	2.45
D	d	2.2857	1	0	1.7143	1.7143
E	e	3.0469	1.5	0	1.2031	1.2031
F	f	3.6667	2	0	0.8333	0.8333
G	g	4.075	2.5	0	0.525	0.525
H	h	4.3909	3	0	0.3091	0.3091
I	i	4.6375	3.5	0	0.1625	0.1625
J	j	4.8308	4	0	0.0692	0.0692
K	k	5	5	0	0	0
L	k	6	5	1	0	1
M	m	6.9222	5.2	0	0.0778	0.0778
N	n	7.3629	5.3	0	0.1371	0.1371
O	o	7.7589	5.5	0	0.2411	0.2411
P	p	8.1237	5.8	0	0.3763	0.3763
Q	q	8.4816	6.2	0	0.5184	0.5184
R	r	8.8574	6.7	0	0.6426	0.6426
S	s	9.271	7.3	0	0.729	0.729

Finally we computed the adjusted score θ^{SAS} and the projected input and output as exhibited in Table 6 while Figure 7 displays them graphically.

Table 6: Scale&cluster-adjusted score and projected input and output

DMU	Adjusted-Score θ^{SAS}	Projected x $(\overline{\overline{x}})$	Projected y $(\overline{\overline{y}})$	Cluster
A	0.4375	0.875	0.5	a
B	0.2917	0.875	0.5	a
C	0.3	1.05	0.6	a
D	0.5714	2.2857	1	d
E	0.7169	3.0469	1.5	e
F	0.8148	3.6667	2	f
G	0.8859	4.075	2.5	g
H	0.9342	4.3909	3	h
I	0.9661	4.6375	3.5	i
J	0.9859	4.8308	4	j
K	1	5	5	k
L	0.8333	6	5	k
M	0.9889	6.9222	5.2	m
N	0.9817	7.3629	5.3	n
O	0.9699	7.7589	5.5	o
P	0.9557	8.1237	5.8	p
Q	0.9424	8.4816	6.2	q
R	0.9324	8.8574	6.7	r
S	0.9271	9.271	7.3	s

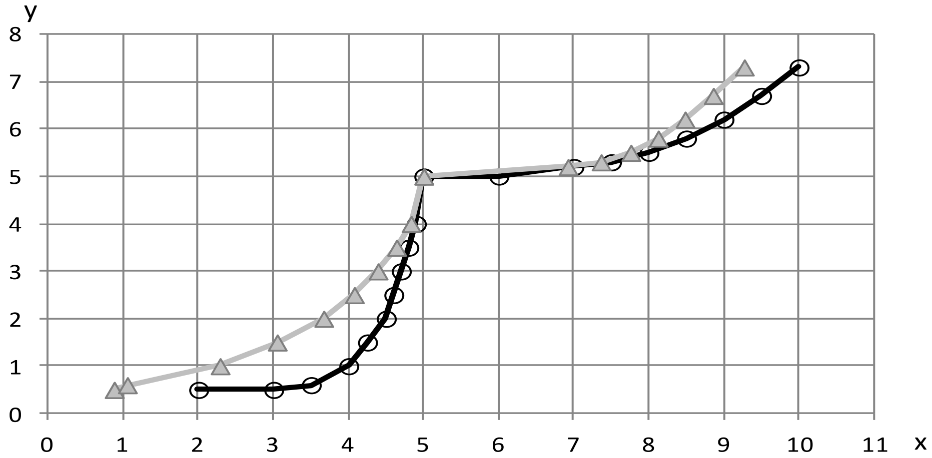

Figure 7: Projection (▲) and data (○)

We compare input-oriented CRS, VRS and SAS scores in Table 7 and Figure 8. Adjusted scores (SAS) of DMUs E to J and M to Q have larger than those of VRS model. This reflects non-convex characteristics of data set.

Table 7: Comparison of three scores

DMU	CRS-I	VRS-I	SAS-I
A	0.25	1	0.4375
B	0.1667	0.6667	0.2917
C	0.1714	0.5905	0.3
D	0.25	0.5833	0.5714
E	0.3529	0.6275	0.7169
F	0.4444	0.6667	0.8148
G	0.5435	0.7246	0.8859
H	0.6383	0.7801	0.9342
I	0.7292	0.8333	0.9661
J	0.8163	0.8844	0.9859

DMU	CRS-I	VRS-I	SAS-I
K	1	1	1
L	0.8333	0.8333	0.8333
M	0.7429	0.7764	0.9889
N	0.7067	0.7536	0.9817
O	0.6875	0.7609	0.9699
P	0.6824	0.7928	0.9557
Q	0.6889	0.8454	0.9424
R	0.7053	0.9153	0.9324
S	0.73	1	0.9271

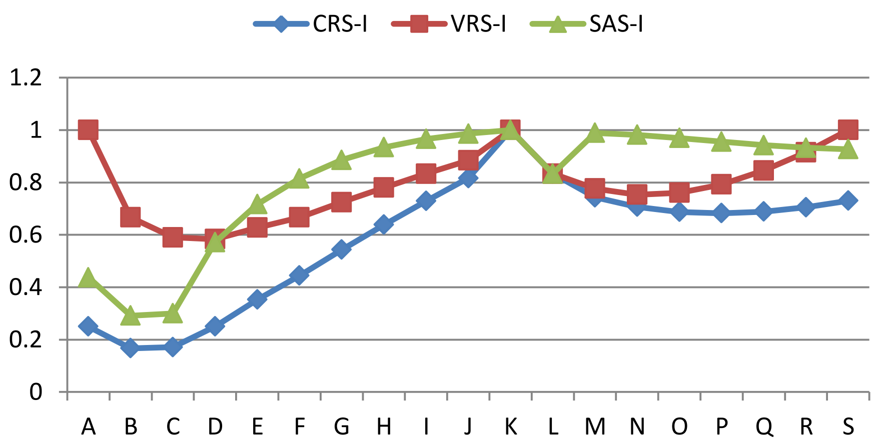

Figure 8: Comparison of three scores

5.2. *Example 2*

Table 8 and Figure 9 exhibit data for Example 2. These DMUs display a typical S-shaped curve.

Table 8: Example 2

DMU	(I)x	(O)y	Cluster
A	2	1	a
B	3	1.2	a
C	4	2	c
D	4.5	3	d
E	5	5	e
F	6	5.8	e
G	7	6.3	g
H	8	6.7	h
I	9	6.9	i
J	10	7	j

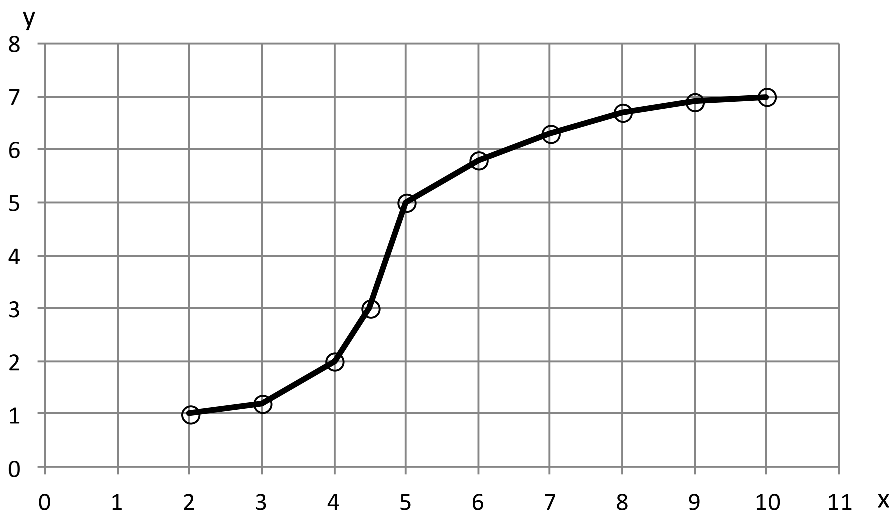

Figure 9: Plot of Example 2

Table 9 and Figure 10 summarize the results of the above procedures. The projected frontiers are a mixture of non-convex and convex parts.

Table 9: Results of Example 2

DMU	CRS-I	VRS-I	SAS-I	$\overset{=}{(x)}$	$\overset{=}{(y)}$	Total slacks	Scale-dependent slacks	In-cluster slacks
A	0.5	1	0.75	1.5	1	0.5	0.5	0
B	0.4	0.7167	0.6	1.8	1.2	1.2	0.7953	0.4047
C	0.5	0.6875	0.8636	3.4545	2	0.5455	0.5455	0
D	0.6667	0.7778	0.9524	4.2857	3	0.2143	0.2143	0
E	1	1	1	5	5	0	0	0
F	0.9667	1	0.9989	5.9933	5.8	0.0067	0.0067	0
G	0.9	1	0.99	6.93	6.3	0.07	0.07	0
H	0.8375	1	0.9736	7.7888	6.7	0.2112	0.2112	0
I	0.7667	1	0.9456	8.51	6.9	0.49	0.49	0
J	0.7	1	0.91	9.1	7	0.9	0.9	0

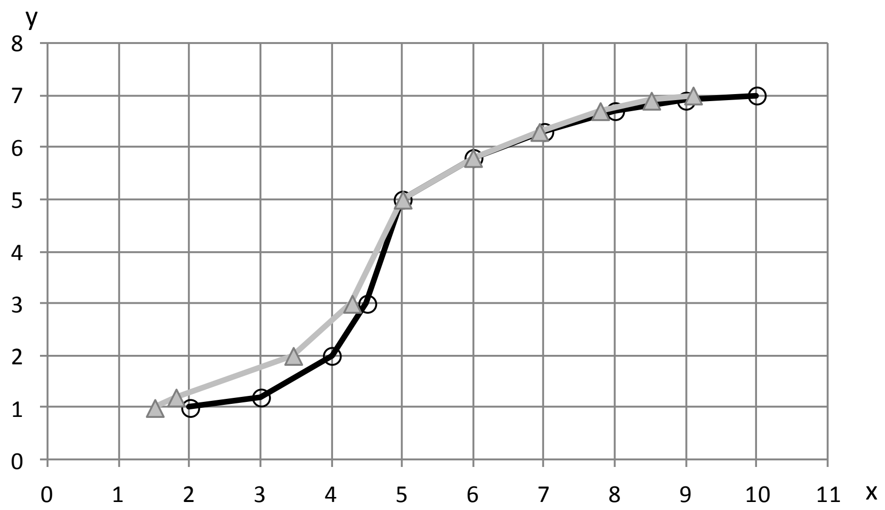

Figure 10: Projection (▲) and data (○)

Figure 11 displays comparison of three scores. At DMUs C and D, Adjusted-scores are larger than VRS scores. This reflects non-convex characteristics of the data set.

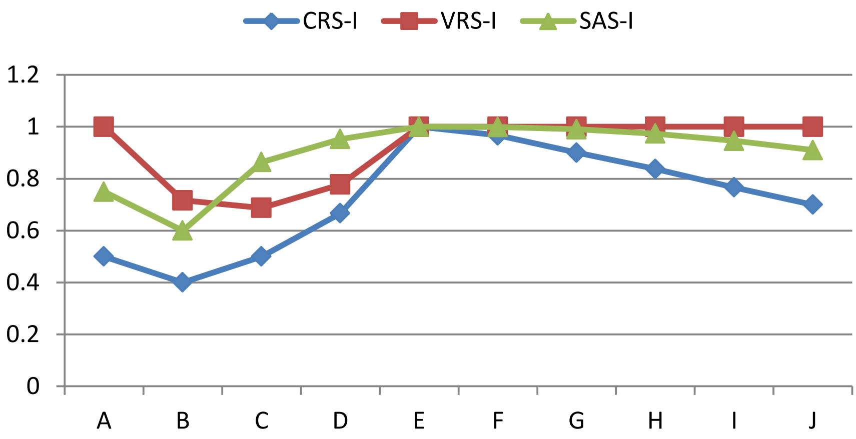

Figure 11: Comparison of three scores

6. Scale-dependent data set and scale elasticity

So far we have discussed the efficiency score issue of our proposed scheme. In this section we deal with the scale elasticity issue. Many papers have discussed this subject under the globally convex frontier assumption. See Banker and Thrall (1992), Banker et al. (2004), Färe and Primond (1995), Førsund and Hjalmarsson (2004a, 2004b), Olsen and Petersen (2013), Podinovski (2004), Kousmanen (2001) among others. However, in case of non-convex frontiers, we believe there is room for further research on this subject. Based on the decomposition of CRS slacks mentioned in Section 2, we develop a new scale elasticity which can cope with non-convex frontiers.

6.1. Scale-dependent data set

We delete or add scale-independent slacks from the data, and thus define the scale-dependent data set $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$.

$$\begin{aligned} &\text{Scale-dependent input} \quad \hat{x}_o = x_o - \sigma_o s_o^{-*} \\ &\text{Scale-dependent output} \; \hat{y}_o = y_o + \sigma_o s_o^{+*} \end{aligned} \tag{17}$$

Figure 12 illustrates an example.

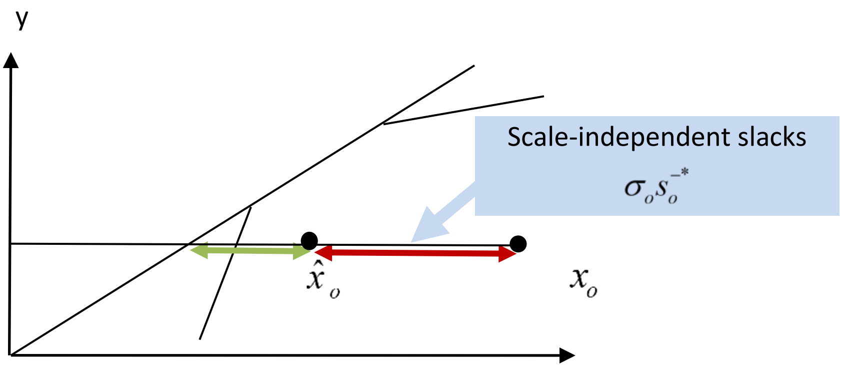

Figure 12: Scale-dependent input

We first project $(\hat{\mathbf{x}}_o, \hat{\mathbf{y}}_o)$ onto the VRS frontier of $(\hat{\mathbf{X}}, \hat{\mathbf{Y}})$ in the same cluster. Thus, we denote them $(\hat{\mathbf{x}}_o^{\mathrm{Proj}}, \hat{\mathbf{y}}_o^{\mathrm{Proj}})$:

$$(\hat{x}_o, \hat{y}_o) \rightarrow (\hat{x}_o^{\mathrm{Proj}}, \hat{y}_o^{\mathrm{Proj}}). \tag{18}$$

6.2. Scale elasticity

The scale elasticity or degree of scale economies is defined as the ratio of marginal product to average product. In a single input/output case, if the output y is produced by the input x, we define the scale elasticity by

$$\varepsilon = \frac{dy}{dx} \Big/ \frac{y}{x}. \tag{19}$$

In the multiple input-output environments, it is determined by solving linear programs related to the supporting hyperplane at the respective efficient point. See Cooper et al. (2007, pp. 147-149) for details.

The production set $(\hat{\mathbf{X}}^{\mathrm{Proj}}, \hat{\mathbf{Y}}^{\mathrm{Proj}})$ defined above has convex frontiers at least within each cluster, we can find a supporting hyperplane at $(\hat{\mathbf{x}}_o^{\mathrm{Proj}}, \hat{\mathbf{y}}_o^{\mathrm{Proj}})$ that supports all projected DMUs in the cluster and has the minimum deviation t from them. This scheme can be formulated as follows:

$$
\begin{aligned}
&\min t \\
&\text{subject to} \\
&\mathbf{v}\hat{\mathbf{x}}_o^{\text{Proj}} = 1 \\
&\mathbf{u}\hat{\mathbf{y}}_o^{\text{Proj}} - u_0 = 1 \\
&-\mathbf{v}\hat{\mathbf{x}}_j^{\text{Proj}} + \mathbf{u}\hat{\mathbf{y}}_j^{\text{Proj}} - u_0 + w_j = 0 \ (\forall j : \text{Cluster}(j) = \text{Cluster}(o)) \\
&-w_j + t \geq 0 \ (\forall j : \text{Cluster}(j) = \text{Cluster}(o)) \\
&\mathbf{v} \geq \mathbf{0}, \mathbf{u} \geq \mathbf{0}, w_j \geq 0 (\forall j), t \geq 0 : u_0 \ \text{free in sign.}
\end{aligned}
\tag{20}
$$

Let the optimal u_0 be u_0^*. We define the scale elasticity of DMU ($\mathbf{x}_o$,$\mathbf{y}_o$) by:

$$
\textbf{Scale Elasticity } \varepsilon_o = \frac{1}{1-u_0^*}. \tag{21}
$$

If u_o^* is not uniquely determined, we check its min and max while keeping *t* at the optimum.

The reason why we apply the above scheme is as follows.

(1) Conventional methods assume a global convex production possibility set for identifying RTS characteristics of each DMU. However, as we observed, the data set not always exhibits convexity. Moreover, the RTS property is a local one, but not global, as the formula (19) indicates. Hence, we discuss this issue within the cluster the DMU belongs to, after deleting the scale-independent slacks.

(2) Conventional methods usually find multiple optimal values of u_0^* and there is a big gap between its min and max. The scale elasticity ε_o defined above remains between the min and max, but has much small allowance.

7. An empirical study

In this section we apply our scheme to a data set comprising 37 Japanese National Universities with the faculty of medicine.

7.1. Data

Table 10 exhibits the data set of Japanese National Universities with the faculty of medicine at the year 2008 (Report by Council for Science and Technology Policy, Japanese Government, 2009). We chose two inputs: (I) Subsidy (unit: one million Japanese yen) and (I) No. of faculty, and three outputs: (O) No. of publication, (O) No. of JSPS (Japan Society for Promotion of Sciences) fund and (O) No. of funded research. We classified them into four clusters: A, B, C and D depending on the sum of No. of JSPS fund and No. of funded research. Cluster A is defined as the set of universities with the sum larger than 2000, Cluster B between 2000 and 1000, Cluster C between 1000 and 500, and Cluster D less than 500.

Table 10: Data set

University	(I)Subsidy	(I)Faculty	(O)Publication	(O) JSPS fund	(O) No. of funded res.	Cluster
A1	96174	4549	6359	2896	2280	A
A2	60868	3562	4776	2304	1504	A
A3	50717	2619	3786	1952	1382	A
A4	50615	2877	4009	1941	1357	A
A5	42398	2207	2605	1396	1186	A
A6	41014	2086	2560	1310	922	A
A7	35985	1792	2443	1351	796	A
B1	48106	1667	1549	911	507	B
B2	28896	1814	1362	811	543	B
B3	22898	1567	1089	751	401	B
B4	18245	1303	1143	606	453	B
B5	18255	1505	1264	606	430	B
C1	19200	1129	803	537	314	C
C2	17569	1010	722	446	302	C
C3	20467	1224	706	428	317	C
C4	16124	1151	582	309	418	C
C5	14515	867	643	351	321	C
C6	17154	1084	685	378	284	C
C7	13196	898	481	325	329	C
C8	12357	830	446	242	357	C
C9	14850	799	628	266	319	C
C10	13138	855	576	353	228	C
C11	16884	1121	531	311	265	C
C12	14589	970	562	277	274	C
C13	14436	976	550	311	229	C
D1	10631	629	293	199	231	D
D2	11319	795	465	190	233	D
D3	10202	657	300	170	240	D
D4	10953	668	311	184	191	D
D5	13017	859	382	201	159	D
D6	11355	775	339	191	156	D
D7	11522	779	391	162	171	D
D8	10637	785	287	174	142	D
D9	8936	656	267	157	153	D
D10	11054	692	343	158	134	D
D11	10888	749	323	157	132	D
D12	10686	645	254	152	135	D

Figure 13 plots 47 universities regarding no. of faculty (input) and no. of publication (output). Globally non-convex characteristics are observed. Especially between big seven universities (A) and other universities (B, C and D), there is a gap. We can see similar gaps among other inputs vs. outputs.

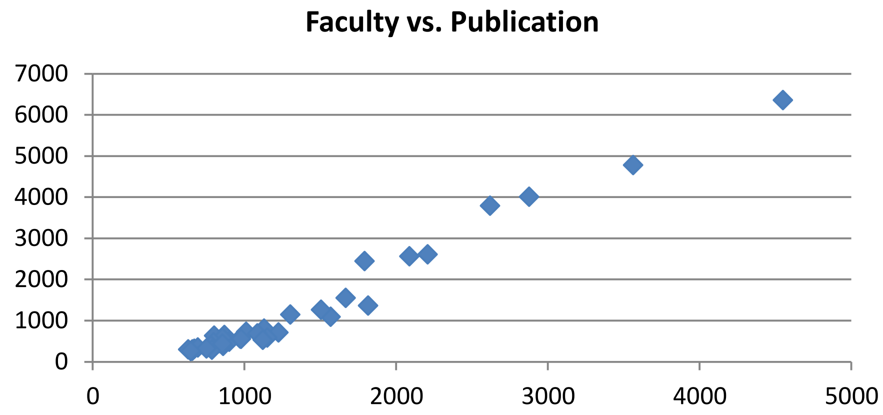

Figure 13: Plot of no. of faculty (horizontal) vs. no. of publication (vertical)

7.2. Adjusted score

Table 11 compares the three scores and Figure 14 displays them graphically.

Table 11: Comparisons of CRS, VRS and SAS (Adjusted score)

DMU	CRS-I	VRS-I	SAS-I	DMU	CRS-I	VRS-I	SAS-I	DMU	CRS-I	VRS-I	SAS-I
A1	0.9246	1	0.9943	C1	0.6824	0.9003	0.9232	D1	0.7301	1	0.9272
A2	0.9764	1	0.9994	C2	0.626	0.8921	0.8885	D2	0.6406	0.9857	0.8742
A3	1	1	1	C3	0.5265	0.7342	0.7287	D3	0.7604	1	0.9426
A4	1	1	1	C4	0.8013	0.8563	0.9872	D4	0.5777	0.9514	0.8033
A5	1	1	1	C5	0.7398	0.9713	0.938	D5	0.394	0.814	0.6426
A6	0.8415	0.9036	0.9891	C6	0.5478	0.8149	0.769	D6	0.4349	0.8796	0.6904
A7	1	1	1	C7	0.7865	0.9994	0.9545	D7	0.4713	0.916	0.7009
B1	0.6126	0.6776	0.9628	C8	1	1	1	D8	0.4089	0.8646	0.6481
B2	0.6645	0.7642	0.8576	C9	0.7554	1	0.9402	D9	0.523	1	0.7725
B3	0.7476	0.8759	0.963	C10	0.626	1	0.8601	D10	0.4029	0.9521	0.6556
B4	0.7794	1	0.9513	C11	0.5005	0.7255	0.6506	D11	0.3847	0.8991	0.6162
B5	0.7395	1	0.9321	C12	0.5985	0.8543	0.7641	D12	0.4206	0.9504	0.6381
				C13	0.5107	0.843	0.7192				

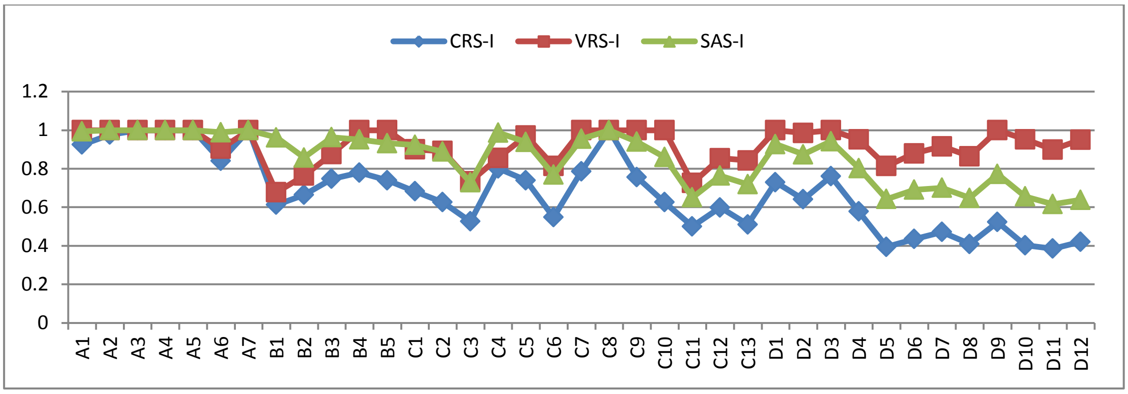

Figure 14: Comparisons of three scores

The SAS of B1, B2 and B3 are remarkably larger than those of VRS, demonstrating the non-convex structure of the data set. Universities in Cluster A are judged almost efficient by adjusted scores. Table 12 summarizes averages of CRS, VRS and SAS for each cluster. For Cluster A universities, gaps among three scores are small and have the highest marks in each model. For Cluster B universities, the average SAS is larger than the average of VRS scores. This indicates the existence of non-convex frontiers around B sized universities. For Cluster C universities, discrepancy between CRS and VRS comes large, and the average of SAS is between them, closer to VRS. For Cluster D universities, the discrepancy comes largest indicating the smallest scale-efficiency. Adjusted scores position around the middle of CRS and VRS. Average SAS decreases monotonically from A to D.

Table 12: Average scores

Cluster	CRS-I	VRS-I	SAS-I
A	0.9632	0.9862	0.9975
B	0.7087	0.8635	0.9334
C	0.6693	0.8916	0.8556
D	0.5124	0.9344	0.7426

7.3. Scale elasticity

Table 14 reports the scale elasticity ε computed by the formula (26).

Table 14: Scale elasticity

DMU	Scale El.	DMU	Scale El.	DMU	Scale El.	DMU	Scale El.
A1	0.961	B1	1.1522	C1	1.137	D1	1.6564
A2	0.9954	B2	1.0915	C2	1.422	D2	1.0532
A3	1.0267	B3	1.1965	C3	1.296	D3	1.7399
A4	1.0299	B4	1.3262	C4	1.152	D4	3.1328
A5	1.0525	B5	1.2003	C5	1.416	D5	1.9453
A6	1.051			C6	1.33	D6	2.034
A7	1.0669			C7	1.197	D7	1.9234
				C8	1.139	D8	3.5783
				C9	1.311	D9	2.1912
				C10	1.56	D10	2.0527
				C11	2.043	D11	2.1179
				C12	2.02	D12	2.1913
				C13	1.56		
Ave.	1.0262	Ave.	1.1933	Ave.	1.429	Ave.	1.642
Max	1.0669	Max	1.3262	Max	2.043	Max	1.9736
Min	0.961	Min	1.0915	Min	1.137	Min	0.6433
StDev	0.0369	StDev	0.0863	StDev	0.303	StDev	0.4143

We observe that for Cluster A universities the scale elasticity is almost unity with the max 1.0669 and min 0.961. This cluster exhibits constant returns-to-scale. Clusters B, C and D universities have elasticity higher than unity and the averages are increasing in this order. They have increasing returns-to-scale characteristics.

8. The radial model case

In this section, we apply the above approaches to the radial DEA models.

8.1. CCR and BCC models

Throughout this section, we utilize the input-oriented radial measures: CCR (Charnes-Cooper-Rhodes (1978)) and BCC (Banker-Charnes-Cooper (1984)) models, for the efficiency evaluation of each DMU (x_o, y_o) $(o = 1, \ldots, n)$ as follows:

$$\begin{aligned} &[\text{CCR}]\ \theta_o^{CCR} = \min \theta \\ &\quad \text{subject to} \\ &\quad \mathbf{X}\boldsymbol{\lambda} \leq \theta \mathbf{x}_o \\ &\quad \mathbf{Y}\boldsymbol{\lambda} \geq \mathbf{y}_o \\ &\quad \boldsymbol{\lambda} \geq \mathbf{0},\ \theta : \text{free}. \end{aligned} \tag{22}$$

$$
\begin{aligned}
&[\text{BCC}]\ \ \theta_o^{BCC} = \min \theta \\
&\quad \text{subject to} \\
&\quad \mathbf{X}\boldsymbol{\lambda} \le \theta \mathbf{x}_o \\
&\quad \mathbf{Y}\boldsymbol{\lambda} \ge \mathbf{y}_o \\
&\quad \mathbf{e}\boldsymbol{\lambda} = 1 \\
&\quad \boldsymbol{\lambda} \ge \mathbf{0}, \theta : \text{free},
\end{aligned}
\tag{23}
$$

where $\boldsymbol{\lambda} \in R^n$ is the intensity vector.

Although we present our model in the input-oriented radial model, we can develop the model in the output-oriented radial model as well.

We define the scale-efficiency (σ_o) of DMU_o by

$$
\sigma_o = \frac{\theta_o^{CCR}}{\theta_o^{BCC}}. \tag{24}
$$

8.2. Decomposition of slacks

We decompose CRS score into scale-independent and –dependent parts as follows:

The radial input-slacks can be defined as

$$
\mathbf{s}_o^- = (1-\theta_o^{CCR})\mathbf{x}_o \in R^m. \tag{25}
$$

We decompose the radial input-slacks into scale-dependent and scale-independent slacks as:

$$
\mathbf{s}_o^- = (1-\sigma_o)\mathbf{s}_o^- + \sigma_o \mathbf{s}_o^- \tag{26}
$$

$$
\begin{aligned}
&\text{Scale-dependent input slacks } \mathbf{s}_o^{ScaleDep-} = (1-\sigma_o)\mathbf{s}_o^- = (1-\sigma_o)(1-\theta_o^{CCR})\mathbf{x}_o \\
&\text{Scale-independent input slacks } \mathbf{s}_o^{ScaleIndep-} = \sigma_o \mathbf{s}_o^- = \sigma_o(1-\theta_o^{CCR})\mathbf{x}_o
\end{aligned}
\tag{27}
$$

8.3. Scale-adjusted input and output

We define scale-adjusted input $\bar{\mathbf{x}}_o$ and output $\bar{\mathbf{y}}_o$ by

$$
\begin{aligned}
\bar{\mathbf{x}}_o &= \mathbf{x}_o - \mathbf{s}_o^{ScaleDep-} = (\sigma_o + \theta_o^{CCR} - \sigma_o \theta_o^{CCR})\mathbf{x}_o \\
\bar{\mathbf{y}}_o &= \mathbf{y}_o.
\end{aligned}
\tag{28}
$$

[Definition 1] (Scale-adjusted score)

We define *scale-adjusted score* by

$$\theta_o^{scale} = \sigma_o + \theta_o^{CCR} - \sigma_o \theta_o^{CCR}. \tag{29}$$

$\bar{\mathbf{x}}_o$ is the scale accounted (free) input.

We have the following propositions.

[Proposition 5]

$$1 \geq \sigma_o + \theta_o^{CCR} - \sigma_o \theta_o^{CCR} \geq \max(\theta_o^{CCR}, \sigma_o) \tag{30}$$

[Proposition 6]

$$\sigma_o + \theta_o^{CCR} - \sigma_o \theta_o^{CCR} = 1 \text{ if and only if } \sigma_o = 1. \tag{31}$$

Proofs are in Appendix A.

8.4. In-cluster issue: Scale&cluster-adjusted score (SAS)

In this section we introduce the cluster of DMUs and define the scale&cluster-adjusted score (SAS).

We classify DMUs into several clusters depending on their characteristics. We denote the name of cluster DMU$_j$ by Cluster(j) $(j = 1, \ldots, n)$.

8.5. Solving the CCR model in the same cluster

We solve the input oriented CCR model for each DMU $(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$ $(o = 1, \ldots, n)$ referring to the $(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ in the same Cluster (o) which can be formulated as follows:

$$\begin{aligned}
&\theta_o^{cl*} = \min \theta_o^{cl} \\
&\text{subject to} \\
&\bar{\mathbf{X}}\boldsymbol{\mu} - \theta_o^{cl} \bar{\mathbf{x}}_o \leq \mathbf{0} \\
&\bar{\mathbf{Y}}\boldsymbol{\mu} \geq \bar{\mathbf{y}}_o \\
&\mu_j = 0 \quad (\forall j : \text{Cluster}(j) \neq \text{Cluster}(o)) \\
&\boldsymbol{\mu} \geq \mathbf{0}, \theta_o^{cl} : \text{free.}
\end{aligned} \tag{32}$$

Scale&cluster adjusted data (projection) $(\bar{\bar{\mathbf{x}}}_o, \bar{\bar{\mathbf{y}}}_o)$ is defined by:

Scale&cluster-adjusted input (Projected Input)

$$\bar{\bar{\mathbf{x}}}_o = \theta_o^{cl*} \bar{\mathbf{x}}_o = \theta_o^{cl*} (\sigma_o + \theta_o^{CCR} - \sigma_o \theta_o^{CCR}) \mathbf{x}_o$$

Output

$$\bar{\bar{\mathbf{y}}}_o = \mathbf{y}_o. \tag{33}$$

[Definition 2] (In-cluster score)

We define θ_o^{cl*} as in-cluster score.

Up to this point, we deleted scale demerits and in-cluster slacks from the data set. Thus, we have obtained a scale free and in-cluster slacks free (projected) data set $(\bar{\bar{\mathbf{X}}}, \bar{\bar{\mathbf{Y}}})$.

8.6. Scale&cluster-adjusted Score (SAS)

[Definition 3] (Scale&cluster-adjusted score)

In the input-oriented case, the scale&cluster-adjusted score (SAS) is defined by

$$\text{Scale\&cluster-adjusted score (SAS)} \quad \theta_o^{SAS} = \theta_o^{cl*} (\sigma_o + \theta_o^{CCR} - \sigma_o \theta_o^{CCR}). \tag{34}$$

Similarly to Propositions 1 to 4, we have the followings.

[Proposition 7] The scale-cluster adjusted score (SAS) is not less than the CCR score.

$$\theta_o^{SAS} \geq \theta_o^{CCR}. \tag{35}$$

[Proposition 8] If $\theta_o^{CCR} = 1$ then it holds $\theta_o^{SAS} = \theta_o^{CCR}$, but not vice versa.

[Proposition 9] The scale-cluster adjusted score (SAS) is decreasing in the increase of input and in the decrease of output so long as the both DMUs remain in the same cluster.

[Proposition 10] The SAS-projected DMU $(\bar{\bar{\mathbf{x}}}_o, \bar{\bar{\mathbf{y}}}_o)$ is radially efficient under the SAS model among the DMUs in the cluster it belongs to. It is also CCR and BCC efficient among the DMUs in its cluster.

9. Concluding remarks

We have developed a scale&cluster-adjusted DEA model assuming scale-efficiency and cluster of DMUs. This model can deal with S-shaped frontiers smoothly. The adjusted score (SAS) reflects the inefficiency of DMUs after deleting the inefficiency caused by scale demerits and accounting in-cluster inefficiency. We also propose a new scheme for evaluation of scale elasticity. We applied this model to a data set comprising Japanese universities.

The managerial implications of this study are as follows.

(1) We are free from the big difference in CRS and VRS scores. Hence, use of DEA becomes more convenient and simple.
(2) We need not any statistical tests on the range of the intensity vector $\boldsymbol{\lambda}$.
(3) We can cope with the non-convex frontiers, e.g. S-shaped curve. In such cases, VRS scores are too stringent to the DMUs.

The optimal slacks are not necessarily determined uniquely. In such a case, we can set the "importance level" of input (output) items and can solve the associated linear programs recursively.

Although we presented the scheme in input-oriented form, we can extend it to output-oriented and non-oriented (both-oriented) model.

Future research subjects include studies in alternative scale-efficiency measures other than the CRS/VRS ratio and clustering methods. Extensions to cost, revenue and profit models are also our future research subjects.

References

Avkiran, N.K. (2001) Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. *Socio-Economic Planning Science*, 35, 57-80.

Avkiran, N.K, Tone, K. and Tsutsui, M. (2008). Bridging radial and non-radial measures of efficiency in DEA. *Annals of Operations Research*, 164, 127-138.

Banker, R. D., Charnes, A. and Cooper, W. W. (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, 30, 1078-1092.

Banker, R.D. and Thrall, R. M. (1992) Estimation of returns to scale using data envelopment analysis. *European Journal of Operational Research*, 62, 74-84.

Banker, R. D., Cooper, W. W., Seiford, L. M., Thrall, R. M. and Zhu, J. (2004) Returns to scale in different DEA models. *European Journal of Operational Research*, 154, 345-362.

Bogetoft, P. and Otto, L. (2010) *Benchmarking with DEA, SFA, and R*. Springer.

Charnes A., Cooper W. W. and Rhodes, E. (1978) Measuring the efficiency of decision – making units. *European Journal of Operational Research,* 2, 429-444.

Cooper, W. W., Seiford, L. M. and Tone, K. (2007) *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software.* Spriger.

Dekker, D. and Post, T. (2001) A quasi-concave DEA model with an application for bank branch performance evaluation. *European Journal of Operational Research*, 132, 296-311.

Färe, R. and Primond, D. (1995) *Multi-output Production and Duality: Theory and Application.* Kluwer Academic Press.

Førsund, F.R. and Hjalmarsson, L. (2004) Are all scales optimal in DEA? theory and empirical evidence. *Journal of Productivity Analysis*, 21, 25-48.

Førsund, F.R. and Hjalmarsson, L. (2004) Calculating scale elasticity in DEA models. *Journal of the Operational Research Society*, 55, 1012-1038.

Kousmanen, T. (2001) DEA with efficiency classification preserving conditional convexity. *European Journal of Operational Research*, 132, 326-342.

Olesen, O. B. and Petersen, N. C. (2013) Imposing the Regular Ultra Passum law in DEA models. *Omega: The International Journal of Management Science*, 41, 16–27.

Podinovski, V.V. (2004) Local and global returns to scale in performance measurement. *Journal of the Operational Research Society*, 55, 170-178.

Tone, K. (2001) A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*,130, 498-509.

Appendix A Proof of Propositions

Let us define the production possibility sets $P(\mathbf{X},\mathbf{Y})$ and $P(\overline{\mathbf{X}},\overline{\mathbf{Y}})$ for $(\mathbf{x}_j,\mathbf{y}_j)$ and $(\overline{\mathbf{x}}_j,\overline{\mathbf{y}}_j)$ $(j=1,\ldots,n)$, respectively by

$$\begin{aligned} P(\mathbf{X},\mathbf{Y}) &= \left\{(\mathbf{x},\mathbf{y}) \middle| \mathbf{x} \geq \sum\nolimits_{j=1}^{n} \mathbf{x}_j\lambda_j, \mathbf{0} \leq \mathbf{y} \leq \sum\nolimits_{j=1}^{n} \mathbf{y}_j\lambda_j, \boldsymbol{\lambda} \geq \mathbf{0}\right\} \\ P(\overline{\mathbf{X}},\overline{\mathbf{Y}}) &= \left\{(\overline{\mathbf{x}},\overline{\mathbf{y}}) \middle| \overline{\mathbf{x}} \geq \sum\nolimits_{j=1}^{n} \overline{\mathbf{x}}_j\lambda_j, \mathbf{0} \leq \overline{\mathbf{y}} \leq \sum\nolimits_{j=1}^{n} \overline{\mathbf{y}}_j\lambda_j, \boldsymbol{\lambda} \geq \mathbf{0}\right\}. \end{aligned} \tag{A1}$$

[Lemma 1] $P(\mathbf{X},\mathbf{Y}) = P(\overline{\mathbf{X}},\overline{\mathbf{Y}})$.

Proof . We define the scale&cluster-adjusted DMU $(\overline{\mathbf{x}}_j,\overline{\mathbf{y}}_j)$ $(j=1,\ldots,n)$ by

$$\bar{\mathbf{x}}_j = \mathbf{x}_j - (1-\sigma_j)\mathbf{s}_j^{-*}$$
$$\bar{\mathbf{y}}_j = \mathbf{y}_j + (1-\sigma_j)\mathbf{s}_j^{+*}. \tag{A2}$$

If $\sigma_j = 1$ (DMUj is efficient), then we have $\bar{\mathbf{x}}_j = \mathbf{x}_j$ and $\bar{\mathbf{y}}_j = \mathbf{y}_j$. If $\sigma_j < 1$ (DMUj is inefficient), then

$$\bar{\mathbf{x}}_j = \mathbf{x}_j - (1-\sigma_j)\mathbf{s}_j^{-*} \geq \mathbf{x}_j - \mathbf{s}_j^{-*}$$
$$\bar{\mathbf{y}}_j = \mathbf{y}_j + (1-\sigma_j)\mathbf{s}_j^{+*} \leq \mathbf{y}_j + \mathbf{s}_j^{+*}, \tag{A3}$$

where $(\mathbf{x}_j - \mathbf{s}_j^{-*}, \mathbf{y}_j - \mathbf{s}_j^{+*})$ is the projection of $(\mathbf{x}_j, \mathbf{y}_j)$ onto the P(**X**,**Y**) frontiers. Thus, $(\bar{\mathbf{x}}_j, \bar{\mathbf{y}}_j)$ $(j = 1, \ldots, n)$ belongs to *P*(**X**,**Y**). Hence, efficient frontiers are common to *P*(**X**,**Y**) and $\boldsymbol{P}(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$. Q.E.D.

[Proposition 1] $\theta_o^{SAS} \geq \theta_o^{CRS} \quad (o = 1, \ldots, n).$

Proof. The CRS scores for $(\mathbf{x}_o, \mathbf{y}_o)$ and $(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$ are, respectively, defined by

$$\text{[CRS]}\ \theta_o^{CRS} = \min 1 - \frac{1}{m}\sum_{i=1}^{m} \frac{s_i^-}{x_{io}}$$
$$\text{subject to}$$
$$\mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^- = \mathbf{x}_o \tag{A4}$$
$$\mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^+ = \mathbf{y}_o$$
$$\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}.$$

and

$$\text{[SAS]}\ \theta_o^{SAS} = \min 1 - \frac{1}{m}\sum_{i=1}^{m} \frac{s_i^{cl-} + (1-\sigma_o)s_i^{-*}}{\bar{x}_{io}}$$
$$\text{subject to}$$
$$\bar{\mathbf{X}}\boldsymbol{\mu} + \mathbf{s}^{cl-} = \bar{\mathbf{x}}_o \tag{A5}$$
$$\bar{\mathbf{Y}}\boldsymbol{\mu} - \mathbf{s}^{cl+} = \bar{\mathbf{y}}_o$$
$$\mu_j = 0 \quad (\forall j : \text{Cluster}(j) \neq \text{Cluster}(o))$$
$$\boldsymbol{\mu} \geq \mathbf{0}, \mathbf{s}^{cl-} \geq \mathbf{0}, \mathbf{s}^{cl+} \geq \mathbf{0}.$$

We prove this proposition in two cases.

(Case 1) All DMUs belong to the same cluster.
In this case (A5) comes to:

$$[\text{SAS}] \quad \theta_o^{SAS} = \min 1 - \frac{1}{m}\sum_{i=1}^{m} \frac{t_i^- + (1-\sigma_o)s_o^{-*}}{\bar{x}_{io}}$$

$$\begin{aligned} & \text{subject to} \\ & \bar{\mathbf{X}}\boldsymbol{\lambda} + \mathbf{t}^- = \bar{\mathbf{x}}_o \\ & \bar{\mathbf{Y}}\boldsymbol{\lambda} - \mathbf{t}^+ = \bar{\mathbf{y}}_o \\ & \boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{t}^- \geq \mathbf{0}, \mathbf{t}^+ \geq \mathbf{0}. \end{aligned} \tag{A6}$$

Let $(\boldsymbol{\lambda}^*, \mathbf{t}^{-*}, \mathbf{t}^{+*})$ be an optimal solution for (A5). Since $P(\mathbf{X}, \mathbf{Y}) = P(\bar{\mathbf{X}}, \bar{\mathbf{Y}})$ and both sets have the same efficient DMUs which span $(\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$, we have

$$\begin{aligned} \mathbf{X}\boldsymbol{\lambda}^* + \mathbf{t}^{-*} &= \bar{\mathbf{x}}_o = \mathbf{x}_o - (1-\sigma_o)\mathbf{s}_o^{-*} \\ \mathbf{Y}\boldsymbol{\lambda}^* - \mathbf{t}^{+*} &= \bar{\mathbf{y}}_o = \mathbf{y}_o + (1-\sigma_o)\mathbf{s}_o^{+*} \end{aligned} \tag{A7}$$

Hence, we have

$$\begin{aligned} \mathbf{X}\boldsymbol{\lambda}^* + \mathbf{t}^{-*} + (1-\sigma_o)\mathbf{s}_o^{-*} &= \mathbf{x}_o \\ \mathbf{Y}\boldsymbol{\lambda}^* - \mathbf{t}^{+*} - (1-\sigma_o)\mathbf{s}_o^{+*} &= \mathbf{y}_o. \end{aligned} \tag{A8}$$

This indicates that $(\boldsymbol{\lambda}^*, \mathbf{t}^{-*} + (1-\sigma_o)\mathbf{s}_o^{-*}, \mathbf{t}^{+*} + (1-\sigma_o)\mathbf{s}_o^{+*})$ is feasible for (A4) and hence its objective function value is not less than the optimal value θ_o^{CRS} .

$$\theta_o^{SAS} = 1 - \frac{1}{m}\sum_{i=1}^{m} \frac{t_i^{-*} + (1-\sigma_o)s_{io}^{-*}}{x_{io}} \geq \theta_o^{CRS}. \tag{A9}$$

(Case 2) Multiple clusters exist.
In this case, we have additional constraints to (A6) for the cluster restriction as follows.

$$[\text{SAS}] \quad \theta_o^{SAS} = \min 1 - \frac{1}{m}\sum_{i=1}^{m} \frac{t_i^- + (1-\sigma_o)s_i^{-*}}{\bar{x}_{io}}$$

$$\begin{aligned} & \text{subject to} \\ & \bar{\mathbf{X}}\boldsymbol{\lambda} + \mathbf{t}^- = \bar{\mathbf{x}}_o \\ & \bar{\mathbf{Y}}\boldsymbol{\lambda} - \mathbf{t}^+ = \bar{\mathbf{y}}_o \\ & \lambda_j = 0 \quad (\forall j : \text{Cluster}(j) \neq \text{Cluster}(o)) \\ & \boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{t}^- \geq \mathbf{0}, \mathbf{t}^+ \geq \mathbf{0}. \end{aligned} \tag{A10}$$

Since adding constrains result in an increase in the objective value, it holds that

$$\theta_o^{SAs} \geq \theta_o^{CRS}. \tag{A11}$$

Q.E.D.

[Proposition 2] If $\theta_o^{CRS} = 1$ then it holds $\theta_o^{SAS} = 1$, but not vice versa.

Proof. If $\theta_o^{CRS} = 1$ then, we have $\mathbf{s}_o^{-*} = \mathbf{0}$ and $\mathbf{s}_o^{+*} = \mathbf{0}$. Hence we have Total slacks = 0 and $\theta_o^{SAS} = 1$. The converse is not always true as demonstrated by the example below where all DMUs belong to an independent cluster.

DMU	(I)x	(O)y	Cluster
A	2	2	a
B	4	2	b
C	6	2	c

DMU	CRS-I	SAS-I	Cluster
A	1	1	a
B	0.5	1	b
C	0.3333	1	c

Q.E.D.

[Proposition 3] The scale&cluster-adjusted score (SAS) is decreasing in the increase of input and in the decrease of output so long as the both DMUs remain in the same cluster.

Proof. Let $\left(\mathbf{x}_p, \mathbf{y}_p\right)$ and $\left(\mathbf{x}_q, \mathbf{y}_q\right)$ with $\mathbf{x}_p \le \mathbf{x}_q$ and $\mathbf{y}_p \ge \mathbf{y}_q$ be respectively the original and varied DMUS in the same cluster. Since the projected point of $\left(\mathbf{x}_p, \mathbf{y}_p\right)$ on the SAS frontiers is feasible for $\left(\mathbf{x}_q, \mathbf{y}_q\right)$ and slacks between $\left(\mathbf{x}_q, \mathbf{y}_q\right)$ and the frontier point are larger than the slacks between $\left(\mathbf{x}_p, \mathbf{y}_p\right)$ and the frontier point. We have this proposition. Q.E.D.

[Proposition 4] The projected DMU $(\overset{=}{\mathbf{x}}_o, \overset{=}{\mathbf{y}}_o)$ is efficient under the SAS model among the DMUs in the cluster it belongs to. It is also CRS and VRS efficient among the DMUs in its cluster.

Proof. From the definition of $(\overset{=}{\mathbf{x}}_o, \overset{=}{\mathbf{y}}_o)$ it is SAS efficient. It is also CRS (VRS) efficient in its cluster. Q.E.D.

[Proposition 5]

$$1 \ge \sigma_o + \theta_o^{CCR} - \sigma_o \theta_o^{CCR} \ge \max(\theta_o^{CCR}, \sigma_o) \tag{A12}$$

Proof. $\sigma_o + \theta_o^{CCR} - \sigma_o \theta_o^{CCR} = \sigma_o (1 - \theta_o^{CCR}) + \theta_o^{CCR} = \theta_o^{CCR} (1 - \sigma_o) + \sigma_o \ge \max\left\{\sigma_o, \theta_o^{CCR}\right\}$.

This term is increasing in σ_o and is equal to 1 when $\sigma_o = 1$. Q.E.D.

[Proposition 6]

$$\sigma_o + \theta_o^{CCR} - \sigma_o \theta_o^{CCR} = 1 \text{ if and only if } \sigma_o = 1. \tag{A13}$$

Proof. If $\sigma_o = 1$, it holds $\sigma_o + \theta_o^{CCR} - \sigma_o \theta_o^{CCR} = 1$.

Conversely, if $\sigma_o + \theta_o^{CCR} - \sigma_o \theta_o^{CCR} = 1$, we have $\sigma_o (1 - \theta_o^{CCR}) = 1 - \theta_o^{CCR}$ Hence, if $\theta_o^{CCR} < 1 \Rightarrow \sigma_o = 1$, else if $\theta_o^{CCR} = 1 \Rightarrow \theta_o^{BCC} = 1$ and $\sigma_o = 1$.

Q.E.D.