Journal of the Operations Research
Society of Japan
2009, Vol. 52, No. 2, 76-85

TUNING REGRESSION RESULTS FOR USE IN MULTI-STAGE DATA ADJUSTMENT APPROACH OF DEA

Kaoru Tone
National Graduate Institute for Policy Studies

Miki Tsutsui
Central Research Institute of Electric Power Industry

(Received March 31, 2008; Revised August 1, 2008)

Abstract Data envelopment analysis (DEA) has been a wildly used powerful method to measure efficiencies of decision making units (DMUs). However, DEA efficiency scores are influenced by uncontrollable factors for respective DMUs. Previous studies attempted separating such factors from DEA scores. Fried et al. [4] proposed a multi-stage data adjustment approach using DEA and a regression model, and several studies have followed it, such as Fried et al. [5], Avkiran and Rowlands [1], and so forth. Firstly, we point out shortcomings of the traditional adjustment scheme for combining regression results for use in DEA in the multi-stage approach, and then we propose a new scheme for data adjustment. We demonstrate the effect of this adjustment formula using an electric utility dataset.

Keywords: DEA, regression, data adjustment, multi-stage approach

1. Introduction

Data envelopment analysis (DEA) has been widely utilized for evaluating relative efficiency of organizations with multiple input resources and output products. DEA is a deterministic method which employs mathematical programming techniques. Since the objective organizations, called Decision Making Units (DMUs), may belong to several different operational environments and their data may subject to statistical noise, it is strongly demanded that the true managerial efficiency should be identified after accounting (deleting) the operating environment effects and statistical noise on the data. For this purpose, Fried et al. [4, 5] proposed a multi-stage procedure that combines DEA and a regression model as follows. At the first stage, they employ DEA for finding slacks of each DMU that constitute the elements of inefficiency. At the second stage, they apply regression models to explain these slacks in terms of the operating environment, statistical noise and managerial inefficiency.* Then, they adjust the first-stage (original) dataset by purging the influence of the operating environment and statistical noise at the third stage.† Lastly, they apply DEA to the adjusted dataset at the fourth stage. Hahn [6], Drake et al. [3], Liu and Tone [7] and Avkiran and Rowlands [1] further developed Fried et al. [5] within the non-radial DEA model, i.e., the slacks-based measure (SBM) introduced by Tone [8].

This paper focuses on the data adjustment schemes at the third stage. Firstly, we

*While Fried et al. [4] decomposed slacks into two terms; operating environmental and statistical noise using Tobit model in the regression stage, Fried et al. [5] employed Stochastic Frontier Analysis (SFA) to decompose slacks into three terms; operating environmental, statistical noise, and managerial inefficiency.
†In Fried et al. [5], the adjusting stage was incorporated into the second stage, and accordingly it was a three-stage procedure. In our paper, following Fried et al. [4], we treat the adjusting stage independently as the third stage.

point out irrationality of the previous adjustment formulae in that the adjustments consist of a positive translation using the regressed terms so that the adjusted data should be non-negative, since most DEA models require non-negative dataset. However, this positive translation causes serious bias in the fourth stage DEA scores as we prove in Section 3. Further, we demonstrate this fact using the results of Stochastic Frontier Analysis (SFA) regression model as examples. Then we propose a new procedure in the third stage for tuning regression results for use in the multi-stage data adjustment of DEA. This paper unfolds as follows. In Section 2, we briefly expose the multiple stages in methodology. Readers are recommended to refer to Fried et al. [4, 5] for detailed discussions on the motivation of the multi-stage approach. In Section 3, we will demonstrate the irrationality of adjustment schemes of previous studies which combine the regression results with the original dataset. Then, we propose a new tuning scheme for adjusting the regression results for use in the multi-stage DEA. Comparisons of our proposed scheme with the previous one are presented in Section 4. Concluding remarks follow in Section 5.

2. Multi-stage Data Adjustment Procedure for DEA

The multi-stage data adjustment procedure proposed by Fried et al. [4] consists of four stages. In this section, we explain it stage by stage.

2.1. Initial measurement of slacks by DEA–1st stage

We deal with n DMUs with the input matrix $\mathbf{X} \in R^{m\times n}$ and output matrix $\mathbf{Y} \in R^{s\times n}$, where m and s are numbers of inputs and outputs, respectively. For the target DMU $(\mathbf{x}_o, \mathbf{y}_o)$, where $\mathbf{x}_o \in R^m_+$ and $\mathbf{y}_o \in R^s_+$ are inputs and outputs of the DMU, we express them in terms of $\mathbf{X}$, $\mathbf{Y}$, the intensity vector $\lambda \in R^s_+$, the input slacks $\mathbf{s}^- \in R^m_+$ and the output slacks $\mathbf{s}^+ \in R^s_+$ as follows:

$$\mathbf{x}_o = \mathbf{X}\lambda + \mathbf{s}^-, \quad \mathbf{y}_o = \mathbf{Y}\lambda - \mathbf{s}^+. \tag{2.1}$$

Fried et al. [5] and Avkiran and Rowlands [1] evaluate the input slacks $\mathbf{s}^- \in R^m_+$ and the output slacks $\mathbf{s}^+ \in R^s_+$, which represent inefficiency of DMU $(\mathbf{x}_o, \mathbf{y}_o)$, by means of DEA models. Difference exists in the DEA models utilized as follows.

Fried et al. [5] employs the input-oriented BCC model (Banker et al. [2]):

$$\begin{aligned} & \min\ \theta \\ \text{subject to}\quad & \theta\mathbf{x}_o = \mathbf{X}\lambda + \mathbf{s}^- \\ & \mathbf{y}_o = \mathbf{Y}\lambda - \mathbf{s}^+ \\ & \mathbf{e}\lambda = 1 \\ & \lambda \geq \mathbf{0},\quad \mathbf{s}^- \geq \mathbf{0},\quad \mathbf{s}^+ \geq \mathbf{0}, \end{aligned} \tag{2.2}$$

where $\mathbf{e} \in R^n$ denotes a row vector in which all elements are equal to 1.

Avkiran and Rowlands [1] utilize the non-radial slacks-based model (SBM) introduced by Tone [8]:

$$\begin{aligned} & \min\ \rho = \frac{1 - \frac{1}{m}\sum_{i=1}^{m} \frac{s_i^-}{x_{io}}}{1 + \frac{1}{s}\sum_{r=1}^{s} \frac{s_r^+}{y_{ro}}} \\ \text{subject to}\quad & \mathbf{x}_o = \mathbf{X}\lambda + \mathbf{s}^- \end{aligned}$$

$$\begin{aligned} \mathbf{y}_o &= \mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^+ \\ \mathbf{e}\boldsymbol{\lambda} &= 1 \\ \boldsymbol{\lambda} \geq \mathbf{0}, &\quad \mathbf{s}^- \geq \mathbf{0}, \quad \mathbf{s}^+ \geq \mathbf{0}. \end{aligned} \tag{2.3}$$

Refer to Avkiran and Rowlands [1] for comparisons of these two approaches. We will not go into the details but just denote the slacks obtained by $\mathbf{s}^-$ and $\mathbf{s}^+$.

2.2. Decomposition of slacks using regression model (SFA)–2^{nd} stage

Both of the previous studies regarded these slacks ($\mathbf{s}^-$ and $\mathbf{s}^+$) as the sources of inefficiencies. However, actual performances are likely to be attributable to some combination of managerial inefficiencies, environmental effects and statistical noise. Thus, they tried isolating these three effects using SFA in the second stage.‡ The general function of the SFA regressions is represented in (2.4) below for the case of input slacks.

$$s_{ij}^- = z_j^i\beta^i + v_{ij} + u_{ij}, \quad (i = 1,\ldots,m; j = 1,\ldots,n) \tag{2.4}$$

where s_{ij}^- is the 1^{st} stage slack in the i^{th} input for the j^{th} unit, z_j^i the environmental variables, β^i the parameter vectors for the feasible slack frontier and $v_{ij} + u_{ij}$ the compounded error structure where $v_{ij} \sim N(0, \sigma_{vi}^2)$ represents statistical noise and $u_{ij} \geq 0$ $(\sim N^+(\mu_i, \sigma_{ui}^2))$ represents managerial inefficiency.

In the same manner, the SFA regressions for output slacks are formulated as follows:

$$s_{rj}^+ = z_j^r\beta^r + v_{rj} + u_{rj}. \quad (r = 1,\ldots,s; j = 1,\ldots,n) \tag{2.5}$$

2.3. Adjustments of original data by regression results – 3^{rd} stage (previous studies)

Fried et al. [5] and Avkiran-Rowlands [1] proposed the following adjustment schemes.§

Fried et al. [5] adjust the input data by deleting significant environmental effects and statistical noises as follows:

Input adjustment

$$x_{ij}^A = x_{ij} + [\max_k\{z_k^i\hat{\beta}^i\} - z_j^i\hat{\beta}^i] + [\max_k\{\hat{v}_{ik}\} - \hat{v}_{ij}]. \tag{2.6}$$

Avkiran and Rowlands [1] adjust the output data as follows:

Output adjustment

$$y_{rj}^A = y_{rj} + [z_j^r\hat{\beta}^r - \min_k\{z_k^r\hat{\beta}^r\}] + [\hat{v}_{rj} - \min_k\{\hat{v}_{rk}\}]. \tag{2.7}$$

The role of max and min in the above formulae is to ensure the adjusted data $\{x_{ij}^A\}$ and $\{y_{rj}^A\}$ to be positive, since most DEA models demand the dataset to be positive. This operation is a translation using the SFA results. Actually, in the input adjustment case, let us define $\hat{z}_i = \max_k\{z_k^i\hat{\beta}^i\}$ and $\hat{v}_i = \max_k\{\hat{v}_{ik}\}$. Then $\hat{z}_i$ and $\hat{v}_i$ are fixed (constant) for all DMUs within the input item i. Thus, (2.6) can be written as

$$x_{ij}^A = x_{ij} - z_j^i\hat{\beta}^i - \hat{v}_{ij} + \hat{z}_i + \hat{v}_i. \tag{2.8}$$

As this formula indicates, the original data are uniformly translated by $\hat{z}_i + \hat{v}_i$ for each i. In Section 3, we point out the troubles that this translation induces.

‡Fried et al. [4], Drake et al. [3] and Hahn [6] utilized Tobit as the regression model in the second stage. Compared to Tobit model, SFA has an advantage to identify managerial inefficiency as well as environmental effects and statistical noise.

§Previous studies using Tobit model such as Fried et al. [4], Drake et al. [3] and Hahn [6] also used this formula excepting the error term adjustment.

2.4. Re-running DEA model using the adjusted data – 4^{th} stage

After adjusting all inputs or outputs, they re-run DEA model with adjusted data $\mathbf{x}_o^A$ or $\mathbf{y}_o^A$ for a specific DMU$_o$ $(o = 1, \dots, n)$ and $\mathbf{X}^A = (\mathbf{x}_1^A, \dots, \mathbf{x}_n^A)$ or $\mathbf{Y}^A = (\mathbf{y}_1^A, \dots, \mathbf{y}_n^A)$ instead of original data in formulae (2.2) and (2.3). It can be said that the new efficiency score obtained at this stage reflects the pure managerial efficiency for each DMU.

3. Shortcomings of Previous Adjustments in the Third Stage

In the multi-stage data adjustment procedure proposed by Fried et al. [4], the third stage adjusts the original data using the regression results estimated in the second stage. However, adjusting formulae employed by previous studies, such as Fried et al. [5], Avkiran and Rowlands [1] and so forth, may cause serious bias in the fourth stage results because of a translation by adding a fixed (constant) value. In this section, we demonstrate irrationality of the adjustment scheme using two examples.

3.1. Two DMUs with single input and single output case

The adjustment formulae (2.6) and (2.7) are introduced so that the adjusted values are assured to be positive. This means a translation by adding a positive value to the original data. Now, we investigate how a positive translation affects DEA efficiency scores using a simple example of the input-oriented case. This example deals only with translation issues but not with environmental and noise issues. Table 1 exhibits two DMUs A and B with a single input x and a single output y. We translate the input x by k. Thus, A's input is $1 + k$ while B's is $2 + k$. Figure 1 depicts these shifts from A to A′ and from B to B′. We translate only input values but keep the output values unchanged.

Table 1: A simple example

	Input	Output		Translated Input	Output
	x	y		$x+k$	y
A	1	2	A′	$1+k$	2
B	2	1	B′	$2+k$	1

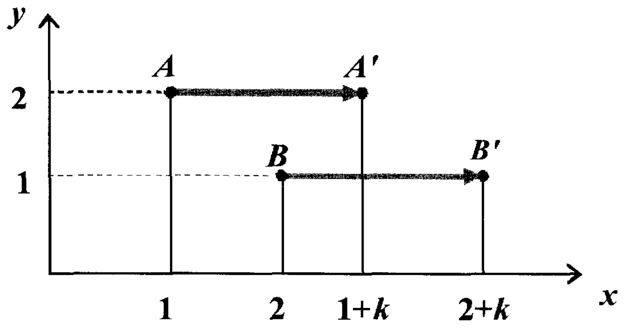

Figure 1: Input translation

In both cases, i.e. the original and the translated cases, A and A′ are efficient, and B and B′ are inefficient compared with A and A′, respectively.

The input-oriented DEA efficiency scores of B′ are calculated in terms of k as follows: Under the constant returns-to-scale assumption (CRS) (CCR-I)

$$\theta_C(k) = \frac{1+k}{2(2+k)}. \tag{3.1}$$

Under the variable returns-to-scale assumption (VRS) (BCC-I)

$$\theta_V(k) = \frac{1+k}{2+k}. \tag{3.2}$$

They are monotone increasing in k and hence the difference in efficiency between A′ and B′ is monotone decreasing in k. Actually, the BCC-I score of B′ tends to unity (that of A′) as k tends to infinity. This simple example demonstrates that the input translation factor k affects the efficiency score significantly. The same holds true for output translation for an output-oriented model. The fact indicates that the adjustment formulae (2.6) and (2.7) suffer from the max and min values included that are translation terms in the respective formula. We notice that this irrationality occurs not only in radial models such as CCR and BCC, but also in non-radial models such as SBM.

The next example will evidence this fact.

3.2. A numerical example

We demonstrate irrationality of the adjustment formula (2.6) using an actual dataset.

3.2.1. Data and statistics

We employed the data from 48 U.S. electric utilities during the years 1990-2001 obtained from "Form No.1" and "Form No.423" published by the Federal Energy Regulatory Commission (FERC) and "Form EIA-860" published by Energy Information Administration (EIA). We count a utility at a certain year as an independent DMU and, after deleting outliers, we obtained 351 utilities as our DMUs. We employed three inputs and one output as follows:

(1) Input

Input 1: Capital Input – The total nameplate capacity of electric power plants measured in Mega Watts (MW)
Input 2: Fuel Input – The consumed fuel converted to British Thermal Units (BTU)
Input 3: Labor Input – The number of employees

(2) Output

Output 1: The generated electric power measured in Mega Watt hours (MWh)

Statistics on the data are displayed in Table 2. They are obtained from the source data divided by some standard of each item.

Table 2: Statistics of the data

	Input 1	Input 2	Input 3	Output 1
Average	0.758	0.796	1.617	1.180
Min.	0.131	0.099	0.135	0.183
Max.	2.268	2.496	7.386	3.437
S.D.	0.515	0.548	1.485	0.787

3.2.2. Model

We applied the following models for each stage.

First stage DEA

We employed the input-oriented SBM under the variable returns-to-scale (VRS) assumption. The results of the 1^{st} stage input-oriented SBM are summarized in Table 3.

Second stage SFA

We applied SFA for the input slacks obtained in the 1^{st} stage SBM. We employed several environmental factors. For the slack of capital input, we employed load factor of the power

Table 3: Results of 1^{st} stage SBM

	Average	Min	Max	S.D.
SBM score	0.719	0.452	1	0.138

plants (LOAD) and commercial customer ratio (CR), which have considerable influence on the capital efficiency.¶ For the fuel input slack, generated power from nuclear, hydraulic and fossil power plants, respectively (NUC, HYD and FOS) were regarded as uncontrollable factors, because power mix affects fuel efficiency and cannot be changed in the short term. In addition, nameplate capacity (NC) was utilized as an explanatory variable in SFA model, which would capture scale effect. However, in this study, we regarded it as a controllable factor for DMUs, and thus, we did not use the estimate of NC for data adjustment. For the slack of labor input, we employed nuclear and hydraulic power ratio (NUCR and HYDR) and NC. We utilized LIMDEP 8.0 for computation. The summary of SFA results is listed in Table 4. For data adjustment, we utilized only significant results.‖

Table 4: Results of 2^{nd} stage SFA

Capital Input Slack

	Coeff.	*t-ratio*	
Const	-0.027	-0.285	
Load	0.0001	0.076	
CR	0.223	3.082	**
Theta	9.422	4.979	**
Sigmav	0.086	87.975	**

Fuel Input Slack

	Coeff.	*t-ratio*	
Const	0.007	0.554	
NUC	-0.233	-3.985	**
HYD	-0.280	-1.341	
FOS	-0.110	-4.122	**
NC	0.162	5.397	**
Theta	9.116	4.589	**
Sigmav	0.065	80.216	**

Labor Input Slack

	Coeff.	*t-ratio*	
Const	-0.133	-0.721	
NUCR	0.357	1.149	
HYDR	4.741	3.623	**
NC	0.306	2.036	*
Theta	2.263	3.376	**
Sigmav	0.593	52.977	**

**: 1% significant level, *: 5% significant level

Third stage data adjustments

We obtained the adjusted inputs using the SFA results by means of the formula (2.6). As mentioned, this formula could be rewritten as (2.8) and the terms $\hat{z}_i = \max_k\{z_k^i\hat{\beta}^i\}$ and $\hat{v}_i = \max_k\{\hat{v}_{ik}\}$ are fixed (constant) for all DMUs within the input i. Hence, the adjustment formula (2.6) becomes to a translation as we denoted in the preceding section.

We record these max terms for each input item in Table 5.

Table 5: The max values

	Input 1	Input 2	Input 3
$\hat{z}_i = \max_k\{z_k^i\hat{\beta}^i\}$	0.114	-0.002	1.399
$\hat{v}_i = \max_k\{\hat{v}_{ik}\}$	0.524	0.575	4.894

Statistics of the adjusted data are summarized in Table 6.

¶In addition, we employed several DMU dummies as explanatory variables in SFA for computational reason. However we did not utilize their estimates for data adjustment. This treatment was conducted in the other input slacks as well.

‖It should be noted that SFA results are obtained by the maximum likelihood estimation, and thus, the estimates may not be uniquely defined because of possibility of other local solutions. Also depending on the choice of explanatory variables, the environmental effects eliminated in the third stage might vary.

Table 6: Statistics of the adjusted data

	Input 1	Input 2	Input 3	Output 1
Average	1.259	1.265	7.329	1.180
Min.	0.779	0.724	6.436	0.183
Max.	3.137	3.089	14.975	3.437
S.D.	0.464	0.512	1.065	0.787

Fourth stage DEA

We applied the input-oriented SBM under variable returns-to-scale assumption to the adjusted dataset. Statistics of the efficiency score are recorded in Table 7.

Table 7: Results of 4^{th} stage SBM using the adjusted data

	Average	Min	Max	S.D.
SBM score	0.985	0.913	1	0.016

Comparisons of Table 3 and Table 7 demonstrate a big change in the average score: from 0.719 to 0.985. Figure 2 compares the distributions of the efficiency scores at the 1^{st} and 4^{th} stage SBM. This increase in the average score might be caused by the adjustment formula (2.6) using the max values for preventing negative input values. The results of the 4^{th} stage SBM almost lost the discriminating power in efficiency evaluation and are unacceptable. Although we described our experiences with the VRS model, we have experienced similar odd results under the constant returns-to-scale (CRS) assumption.

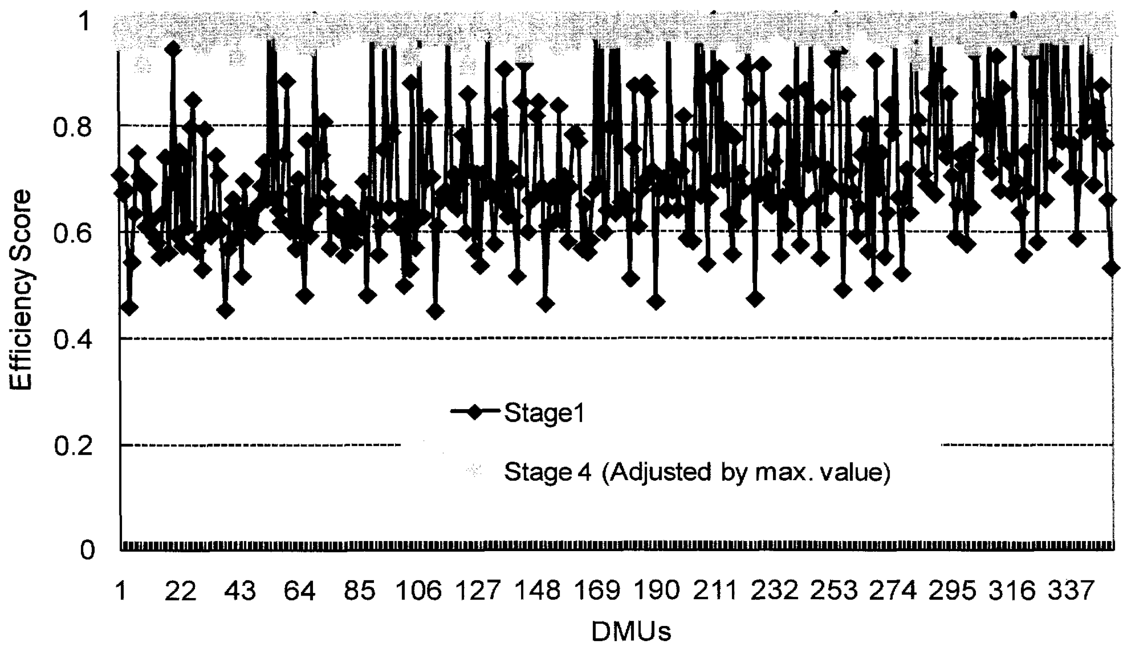

Figure 2: Comparisons of stage 1 and stage 4 efficiency scores

4. A New Tuning Procedure of Regression Results

In this section, we propose a new adjustment scheme.

4.1. Re-adjustments

First, we employ the formula for adjustment with no recourse to max or min as follows:

Input adjustment

$$x_{ij}^{A}=x_{ij}-z_{j}^{i}\hat{\beta}^{i}-\hat{v}_{ij}. \tag{4.1}$$

Output adjustment

$$y_{rj}^{A}=y_{rj}+z_{j}^{r}\hat{\beta}^{r}+\hat{v}_{rj}. \tag{4.2}$$

Then we re-adjust them into x_{ij}^{AA} or y_{rj}^{AA} using the following formulae.

Input Re-adjustment

$$x_{ij}^{AA}=\frac{x_{imax}-x_{imin}}{x_{imax}^{A}-x_{imin}^{A}}(x_{ij}^{A}-x_{imin}^{A})+x_{imin}, \quad (i=1,\ldots,m:j=1,\ldots,n) \tag{4.3}$$

where $x_{imax}=\max_k\{x_{ik}\}, x_{imin}=\min_k\{x_{ik}\}, x_{imax}^{A}=\max_k\{x_{ik}^{A}\}$ and $x_{imin}^{A}=\min_k\{x_{ik}^{A}\}$.

Output Re-adjustment

$$y_{rj}^{AA}=\frac{y_{rmax}-y_{rmin}}{y_{rmax}^{A}-y_{rmin}^{A}}(y_{rj}^{A}-y_{rmin}^{A})+y_{rmin}, \quad (r=1,\ldots,s:j=1,\ldots,n) \tag{4.4}$$

where $y_{rmax}=\max_k\{y_{rk}\}, y_{rmin}=\min_k\{y_{rk}\}, y_{rmax}^{A}=\max_k\{y_{rk}^{A}\}$ and $y_{rmin}^{A}=\min_k\{y_{rk}^{A}\}$.

4.2. Rationale

The proposed re-adjustment scheme has the following properties:

(1) x_{ij}^{AA} increases in x_{ij}^{A}. Thus, the re-adjusted data have the same ranking with the adjusted data. Actually x_{ij}^{AA} is a linear transformation of x_{ij}^{A} with a positive coefficient. The coefficient and the constant term of this linear transformation are constant within the respective input item i.

(2) At x_{imax}^{A}, x_{imax}^{AA} attains the maximum value $x_{imax}^{AA}=x_{imax}$.

(3) At x_{imin}^{A}, x_{imin}^{AA} attains the minimum value $x_{imin}^{AA}=x_{imin}$.

Hence, the re-adjusted dataset $\{x_{ij}^{AA}\}$ remains in the range $[x_{imin},x_{imax}](\forall i)$, and the maximum and minimum values are the same between $\{x_{ij}^{AA}\}$ and $\{x_{ij}\}$.

For the output side, we have the same property: the re-adjusted dataset $\{y_{rj}^{AA}\}$ remains in the range $[y_{rmin},y_{rmax}](\forall r)$, and the maximum and minimum values are the same between $\{y_{rj}^{AA}\}$ and $\{y_{rj}\}$.

These properties are appealing in that they eliminate ambiguity regarding the range of adjusted input and output values that affects the DEA scores significantly as we have shown in the previous examples. Furthermore, when we start the first stage DEA, we usually confirm that the ranges of input and output values are appropriate for the chosen DEA model. (We delete outliers before going into the first stage.) Therefore, it is not odd to keep the ranges status quo and re-evaluate the DEA efficiency score at the fourth stage using the re-adjusted dataset.

4.3. Numerical comparisons

We re-adjust the U.S. electric utility data introduced in Section 3.2 and compare the results. Using the formula (4.1) (but not using the max in (2.6)), we adjusted the input data, and then re-adjusted the data by the formula (4.3). Table 8 displays the statistics of the re-adjusted data. As expected, the min and max values are the same with the original data in Table 2.

The new 4^{th} stage SBM was applied to this re-adjusted dataset and the results are summarized in Table 9.

Table 8: Statistics of the re-adjusted data

	Input 1	Input 2	Input 3	Output 1
Average	0.566	0.648	0.894	1.180
Min	0.131	0.099	0.135	0.183
Max	2.268	2.496	7.386	3.437
S.D.	0.421	0.519	0.904	0.787

Table 9: Results of the new 4^{th} stage SBM using the re-adjusted data

	Average	Min	Max	S.D.
SBM score	0.923	0.681	1.00	0.075

Figure 3 compares the efficiency scores of the 1^{st} and the new 4^{th} stage SBM. The upgrade of the average score from 0.719 (1^{st} stage) to 0.923 (New 4^{th} stage) reflects the effects of environmental factors and statistical noises identified in the 2^{nd} stage SFA. Compared with the Figure 2 which resulted from the adjustments using max value, the new 4^{th} stage results are more acceptable for efficiency evaluations.

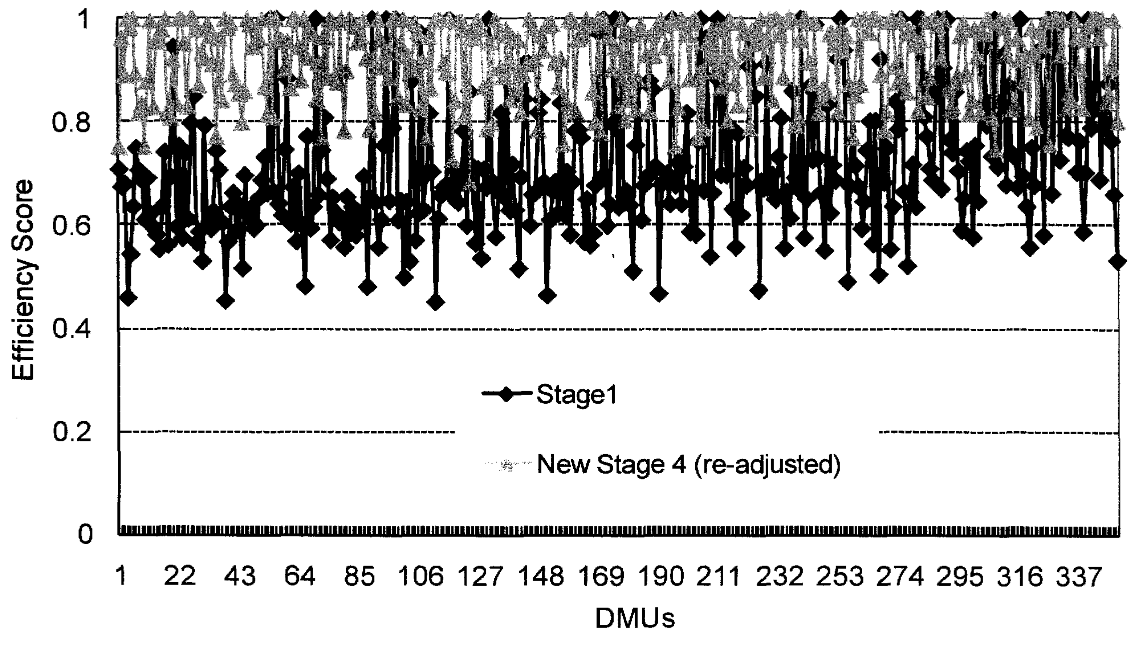

Figure 3: Comparisons of stage 1 and new stage 4 efficiency scores

5. Concluding Remarks

In the DEA studies, many authors have tried to identify the true managerial efficiency after accounting for the operational environment effects and statistical noises on the data. The multi-stage approach proposed by Fried et al. [4, 5] is a remarkable advance on this line. They combined DEA with regression model in the manner that the slacks obtained in the 1^{st} stage DEA was regressed by means of the environmental effects, statistical noises and managerial efficiency** in the data. Then they adjust the original input data using the regression results.

In this paper, after pointing out shortcomings in their data adjustment, we proposed a new adjustment scheme using regression results for use in DEA. This scheme was applied to U.S. electric utilities and proved its superiority over the traditional one. Combining non-parametric DEA with parametric model may arouse several serious problems e.g. selection of distribution type and functional form. The data adjustment problem is an important issue among them. We hope our method serves as a stepping stone to the final resolution.

Acknowledgements

The authors wish to acknowledge the useful comments and suggestions of the two anonymous referees.

References

[1] N.K. Avkiran and T. Rowlands: How to better identify the true managerial performance: State of the art using DEA. *OMEGA*, **36** (2008), 317–324.

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

[3] L. Drake, M.J. Hall and R. Simper: The impact of macroeconomic and regulatory factors on bank efficiency: A non-parametric analysis of Hong Kong's banking system. *Journal of Banking and Finance*, **30** (2006), 1443–1466.

[4] H.O. Fried, S.S. Schmidt and S. Yaisawarng: Incorporating the operating environment into a nonparametric measure of technical efficiency. *Journal of Productivity Analysis*, **12** (1999), 249–267.

[5] H.O. Fried, C.A.K. Lovell, S.S. Schmidt and S. Yaisawarng: Accounting for environmental effects and statistical noise in data envelopment analysis. *Journal of Productivity Analysis*, **17** (2002), 157–174.

[6] F.R. Hahn: Measuring performance - a multiple-stage approach. *WIFO working papers* (2004), No.228.

[7] J. Liu and K. Tone: A multistage method to measure efficiency and its application to Japanese banking industry. *Socio-Economic Planning Sciences*, **42** (2008), 75–91.

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

Miki Tsutsui
Central Research Institute of Electric Power Industry
2-11-1 Iwado Kita, Komae-shi, Tokyo 201-8511, Japan
E-mail: `miki@criepi.denken.or.jp`

**The managerial efficiency term can be obtained in SFA model, while Tobit model does not provide it in the regression stage.