Journal of the Operations Research
Society of Japan
Vol. 41, No. 4, December 1998

USING THE EXCLUSION MODEL FOR DEA COMPUTATION

Akihiro Hashimoto
University of Tsukuba

(Received December 19, 1996; Final March 23, 1998)

Abstract In the original DEA/CCR (Data Envelopment Analysis/Charnes, Cooper and Rhodes) computation with n DMUs (Decision Making Units), we cannot make do with solving n LP (Linear Programming) problems even to judge only whether each DMU is DEA efficient or not in using ordinary LP solvers. This is because we must use two-phase optimization unless we have access to DEA software packages taking non-Archimedean infinitesimals into consideration. We must solve n Phase I LPs for all the n DMUs plus Phase II LPs to see whether DEA inefficient DMUs on the extended frontier are. This paper shows that, through solving nearly n LPs, we can achieve it if we use the DEA exclusion model instead of the standard DEA model, etc. We should note a merit of the DEA exclusion model for reducing DEA computation load as well.

1. Introduction

DEA (Data Envelopment Analysis) measures the relative efficiency of DMUs (Decision Making Units). The original CCR (Charnes, Cooper and Rhodes) model to obtain DEA efficiency score $h^*_{j_0}$, $0 < h^*_{j_0} \le 1$, for target DMU j_0 is expressed as follows, converted into the LP (Linear Programming) formulation (i.e., the *input oriented multiplier form*) [3, 4]:

$$\text{Maximize} \quad h_{j_0} = \sum_{r=1}^{t} u_r y_{rj_0} \tag{1.1a}$$

$$\text{subject to} \quad \sum_{i=1}^{m} v_i x_{ij_0} = 1, \tag{1.1b}$$

$$\sum_{r=1}^{t} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \; j = 1, ..., n, \tag{1.1c}$$

$$u_r, v_i \ge \epsilon, \; r = 1, ..., t, i = 1, ..., m, \tag{1.1d}$$

where y_{rj} = the amount of output r from DMU j; x_{ij} = the amount of input i to DMU j; u_r = the weight given to output r; v_i = the weight given to input i; n = the number of DMUs; t = the number of outputs; m = the number of inputs; ϵ = a positive non-Archimedean infinitesimal; and $h^*_{j_0}$ = the maximum of h_{j_0}.

The dual of problem (1.1) (i.e., the *input oriented envelopment form*) is as follows:

$$\text{Minimize} \quad p_{j_0} = \theta - \epsilon(\sum_{r=1}^{t} s_r^+ + \sum_{i=1}^{m} s_i^-) \tag{1.2a}$$

$$\text{subject to} \quad \sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = y_{rj_0}, \; r = 1, ..., t, \tag{1.2b}$$

$$x_{ij_0}\theta - \sum_{j=1}^{n} x_{ij}\lambda_j - s_i^- = 0, \; i = 1, ..., m, \tag{1.2c}$$

$$\lambda_j, s_r^+, s_i^- \geq 0,\ j = 1, ..., n, r = 1, ..., t, i = 1, ..., m, \quad (1.2d)$$
$$(\theta \text{ unconstrained}),$$

where θ, λ_j, s_r^+, s_i^- = dual variables. Of course, the minimum $p_{j_0}^* = h_{j_0}^*$ at an optimum. We solve this problem for DMU $j_0, j_0 = 1, ..., n$, to get DEA information.

The set of all DMUs is partitioned into the following four subsets: E, E', F and N [5, 6, 10]. Here, E is a set of DEA efficient DMUs with $\theta^* = 1$, each of which is a vertex on the *efficient frontier*; E' is a set of DEA efficient DMUs with $\theta^* = 1$, each of which is on the efficient frontier but not a vertex; F is a set of DEA inefficient DMUs with $\theta^* = 1$, each of which is on the *extended frontier*; and N is a set of DEA inefficient DMUs with $\theta^* < 1$, each of which is not on the efficient nor extended frontier. Letting n_E denote the number of DMUs belonging to subset E, etc., $n = n_E + n_{E'} + n_F + n_N$. We should here note that, in the general DEA analysis, $n_{E'}$ or n_F is considerably small as compared with n_E or n_N (i.e., $n_{E'} \cong n_F \cong 0$). Fig.1 shows the four subsets of DMUs in the case of two inputs and one output for graphical simplicity.

The DEA information that we have to get through DEA computation is various. This paper defines the three levels of DEA information as follows: *Level 1.* DEA efficiency score judging whether DEA efficient or DEA inefficient and the optimal weights u_r^*, v_i^* for each DMU; *Level 2.* Level 1 information plus the *reference set* and *combination coefficients* for each DEA inefficient DMU belonging to subset N or F; and *Level 3.* Level 2 information plus which subset each DMU belongs to. Here, the reference set of a DEA inefficient DMU j_0 is a set of DMUs j being efficient with the weights optimal to the DMU j_0 and the combination coefficients are λ_j^* for j belonging to the reference set. We should note that DMUs belonging to F as well as N cannot be elements of the reference set because they are not DEA efficient. In general DEA applications, Level 3 information would not always be needed. We would like to obtain DEA information of a certain level as occasion demands.

Since DEA is LP-based, it should be available to any potential user or researcher with access to the ordinary LP solvers. However, a typical DEA application requires the solutions of a large number of LPs. We cannot make do with solving n LPs even to get only Level 1 information, because we must solve problem (1.2) using two-phase optimization unless we have access to DEA software packages taking non-Archimedean infinitesimals into consideration. We must solve *Phase I LP*, problem (1.2) with objective "Minimize θ", for all the n DMUs, plus *Phase II LP*, a modification of Phase I LP in which the objective "Minimize θ" is replaced by

$$\text{Maximize} \quad \sigma_{j_0} = \sum_{r=1}^{t} s_r^+ + \sum_{i=1}^{m} s_i^- \quad (1.3)$$

and θ is fixed $\theta = 1$, in order to see whether DEA inefficient DMUs on the extended frontier are.

However, we can get Level 3 information through solving nearly n LPs without access to DEA software packages but using the *DEA exclusion model* instead of the standard DEA model. Andersen and Petersen [2] proposed the DEA exclusion model, which allows DEA efficiency scores to exceed unity (i.e., the super-efficiency [9]) unlike the standard DEA model. It is said that there are at least three different motivations for the exclusion model [1]: i) to discriminate or to rank DEA efficient DMUs [2] (see [7] for an application); ii) to obtain a nontruncated distribution of DEA efficiency to facilitate analysis of DEA score distributions [9]; and iii) to detect outlying DMUs in the comparison set. Besides these motivations, this paper demonstrates that the exclusion model has a merit for reducing DEA computation load as well.

In the next section, we show that Level 1 or 2 information requires nearly $(n + n_E)$ LPs and Level 3 information does nearly $(n + 2n_E)$ LPs in the standard DEA model case. The third section demonstrates that we can get Level 3 information through solving nearly n LPs by using the DEA exclusion model. Further, the exclusion model originally gives DEA information discriminating DEA efficient DMUs, peculiar to this model vs the standard DEA model. Therefore, we propose to always use the DEA exclusion model for DEA computation as one getting more DEA information through less computation than the standard DEA model.

2. DEA computation

We usually get DEA information by going through the following steps using the standard DEA model [5, 6]:

Step 1.1. Solving problem (1.2) with objective "Minimize θ" (i.e., Phase I LP) for each DMU j_0 of the n DMUs, we obtain the optimal solution $\theta^*, \lambda_j^*, s_r^{+*}$ and s_i^{-*}, and the optimal weights u_r^*, v_i^* as shadow prices of constraints (1.2b) and (1.2c). We can here identify subset N comprising n_N DMUs in terms of $\theta^* < 1$. The remaining $(n - n_N)$ DMUs j_0 with $\theta^* = 1$ would contain DMUs belonging to subset E, E' or F. Note that Phase I LP expresses DMU j_0's radial projection point onto the frontier facet (i.e., *reference point*) in terms of a non-negative linear combination of DMUs on the facet and slacks. DMUs j_0 with $\theta^* = 1$ and no slacks $(s_r^{+*} = s_i^{-*} = 0, r = 1, ..., t, i = 1, ..., m)$ at this step might contain DMUs belonging to subset F as well as subset E or E', because Phase I LP for DMU $j_0 \in F$ has multiple optimal extreme-point solutions and the solution with no slacks might happen to be obtained. If there happen to be DMUs j_0 with $\theta^* = 1$ and nonzero slacks $[\sum_{r=1}^{t} s_r^{+*} + \sum_{i=1}^{m} s_i^{-*} \neq (>)0]$, we can here judge that they belong to F. Let n_π be the number of such DMUs, then $n_\pi \leq n_F$, so that n_π would be very small (i.e., $n_\pi \cong 0$). We should note that, for DMU $j_0 \in N$ or F identified at this step, DMUs j of $\lambda_j^* > 0$ might not necessarily be elements of DMU j_0's reference set because they might be DEA inefficient DMUs on the extended frontier (see Step 1.3).

Step 1.2. For each DMU j_0 of the $(n - n_N - n_\pi = n_E + n_{E'} + n_F - n_\pi)$ DMUs with $\theta^* = 1$ and no slacks at Step 1.1, we solve a modification of problem (1.2) in which the objective is replaced by (1.3) and θ is fixed at $\theta = 1$ (i.e., Phase II LP). We can here identify the rest of subset F comprising $(n_F - n_\pi)$ DMUs in terms of the maximum $\sigma_{j0}^* > 0$, and obtain their reference sets and combination coefficients in terms of $\lambda_j^* > 0$ of this step.

Step 1.3. Suppose that the DMUs belonging to subset F (identified at Step 1.1 or 1.2) are included in the set of DMUs j of $\lambda_j^* > 0$ for DMUs $j_0 \in N$ or F of Step 1.1. This means that DMU j_0's reference point is on the extended frontier, and note that the latter case (for DMUs $j_0 \in F$) can also be occurred when plural DMUs belonging to F are on the extended frontier (see Fig.1). Then, for each of such DMUs j_0, we solve a modified form of Phase II LP in which $\theta = 1$ is replaced by $\theta = \theta^*$, and obtain its reference set and combination coefficients in terms of $\lambda_j^* > 0$ of this step. (We can also achieve this by solving a modified form of Phase I LP in which all λ_j for $j \in F$ are fixed at $\lambda_j = 0$.) Letting n_α denote the number of problems to be solved here, n_α would be very small (i.e., $n_\alpha \cong 0$). For each of the remaining $(n_N + n_\pi - n_\alpha)$ DMUs $j_0 \in N$ and F, the DMUs j of $\lambda_j^* > 0$ at Step 1.1 form DMU j_0's reference set.

Step 1.4. For each DMU j_0 of the $(n_E + n_{E'})$ DMUs with $\sigma^*_{j_0} = 0$ at Step 1.2, we solve a modified form of problem (1.2), this time with the objective "Minimize λ_{j_0}" and again with θ fixed at $\theta = 1$. If the minimum $\lambda^*_{j_0} = 0$, then the DMU j_0 belongs to subset E', otherwise it belongs to E.

In this way, to obtain the DEA information, we must solve considerably many LPs. Note that discriminating DMUs belonging to subset F (subset E') at Step 1.2 (Step 1.4) requires nearly n_E LPs though $n_F \cong 0$ ($n_{E'} \cong 0$). We must solve a total of $(n + n_E + n_{E'} + n_F - n_\pi \cong n + n_E)$ LPs to get Level 1 information throughout Steps 1.1-1.2; $(n + n_E + n_{E'} + n_F + n_\alpha - n_\pi \cong n + n_E)$ LPs to get Level 2 information throughout Steps 1.1-1.3; and $(n + 2n_E + 2n_{E'} + n_F + n_\alpha - n_\pi \cong n + 2n_E)$ to get Level 3 information throughout Steps 1.1-1.4. In fact, in a DEA case with 47 DMUs, four inputs and four outputs [8], we could identify subset N comprising 21 DMUs and $n_\pi = 0$ at Step 1.1. Therefore, we had to solve 26 Phase II LPs at Step 1.2, and eventually found no DMUs belonging to F. That is, we solved 73 LPs to get Level 1 information throughout Steps 1.1-1.2 and to get Level 2 information throughout Steps 1.1-1.3 because $n_\alpha = 0$. Further, if we would like to discriminate between subsets E and E', we must solve 26 more problems at Step 1.4, i.e., a total of 99 problems throughout Steps 1.1-1.4.

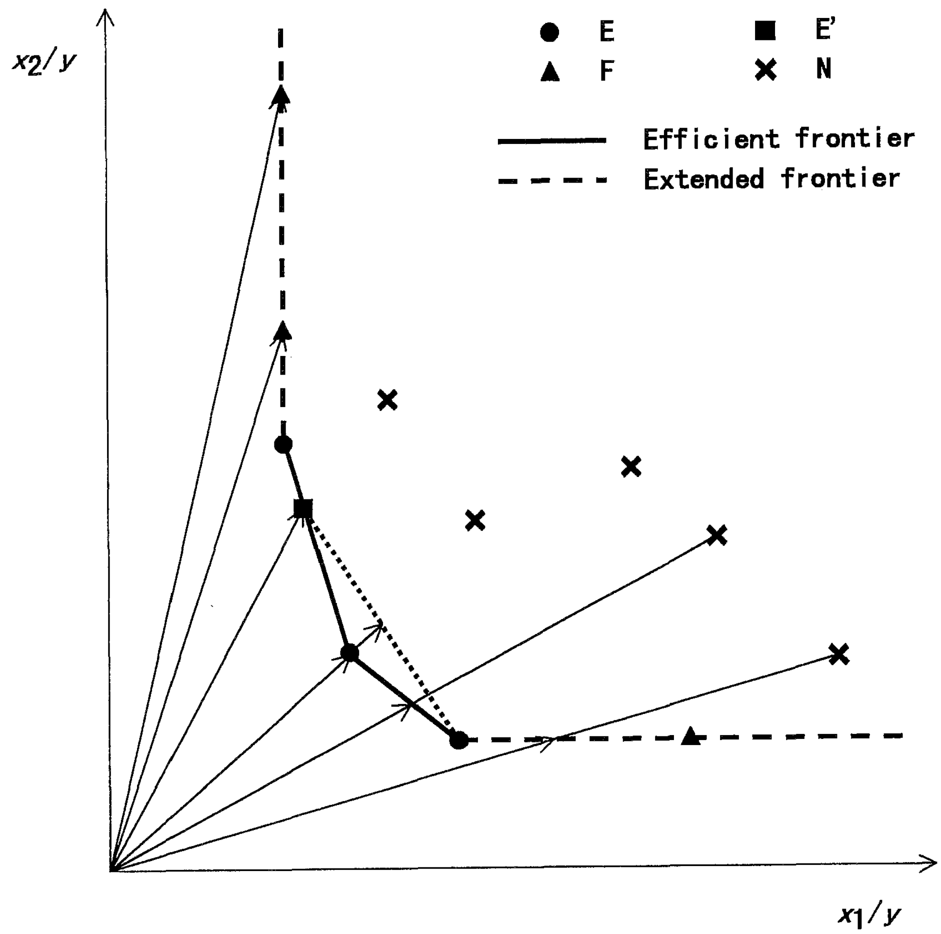

Figure 1. DMUs belonging to the four subsets, reference point vectors onto the frontier facets, and their shifts.

3. Using the DEA exclusion model

In the DEA exclusion model, the DMU being evaluated is excluded from the comparison set [1]. The exclusion model corresponding to model (1.2) is expressed as follows:

$$\text{Minimize} \quad q_{j_0} = \phi - \epsilon\left(\sum_{r=1}^{t} s_r^+ + \sum_{i=1}^{m} s_i^-\right) \tag{3.1a}$$

$$\text{subject to} \quad \sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = y_{rj_0},\ r = 1, ..., t, \tag{3.1b}$$

$$x_{ij_0}\phi - \sum_{j=1}^{n} x_{ij}\lambda_j - s_i^- = 0,\ i = 1, ..., m, \tag{3.1c}$$

$$\lambda_{j_0} = 0, \tag{3.1d}$$

$$\lambda_j, s_r^+, s_i^- \geq 0,\ j = 1, ..., n, r = 1, ..., t, i = 1, ..., m, \tag{3.1e}$$

$$(\phi \text{ unconstrained}),$$

where ϕ = variable used instead of θ in model (1.2). This is different from (1.2) by only constraint (3.1d). Since the region of feasible solutions of the exclusion model is more strictly constrained than that of the standard model (1.2) by (3.1d), the minimum to Phase I LP $\phi^* \geq \theta^*$.

The minimum θ^* for DMU j_0 is the ratio of vector length of DMU j_0's reference point on the frontier facet to that of DMU j_0 itself. Since DMU j_0 belonging to subset E, E' or F is on the frontier, the DMU itself and its reference point are identical. Therefore, $\theta^* = 1$ for DMU $j_0 \in E, E'$ or F and $\theta^* < 1$ for DMU $j_0 \in N$ as mentioned in Sec.1. (See Fig.1.)

On the other hand, the minimum ϕ^* has the similar implication to θ^*. However, excluding the DMU j_0 being evaluated from the comparison set in the exclusion model, the frontier facets and the reference point on them shift for DMU $j_0 \in E$. For DMU $j_0 \in E', F$ or N, the frontier facets do not shift, so that the reference point does not change. Therefore, $\phi^* > \theta^* = 1$ for DMU $j_0 \in E$; $\phi^* = \theta^* = 1$ for DMU $j_0 \in E'$ or F; and $\phi^* = \theta^* < 1$ for DMU $j_0 \in N$ [2].

Using the DEA exclusion model (3.1), we can get the DEA information by going through the following steps:

Step 2.1. Solving problem (3.1) with objective "Minimize ϕ" (i.e., Phase I LP) for each DMU j_0 of the n DMUs, we obtain the optimal solution and shadow prices. We can here identify subset E comprizing n_E DMUs in terms of $\phi^* > 1$ and subset N comprizing n_N DMUs in terms of $\phi^* < 1$. The remaining $(n - n_E - n_N)$ DMUs j_0 with $\phi^* = 1$ would comprise those belonging to subset E' or F. We can here judge that DMUs j_0 with $\phi^* = 1$ and nonzero slacks belong to F. Since the DEA exclusion model does not allow to use DMU j_0 itself in the linear combination expressing its reference point, almost all DMUs $j_0 \in F$ would have nonzero slacks at this step. Let $n_{\pi'}$ be the number of such DMUs, then $n_\pi \leq n_{\pi'} \leq n_F$ and $n_{\pi'} \cong n_F$ (i.e., $n_{\pi'} \cong 0$). But we should note that there is a little possibility that DMUs with $\phi^* = 1$ and no slacks contain not only DMUs $j_0 \in E'$ but also DMUs $j_0 \in F$ in the case where plural DMUs belonging to F are on the extended frontier (see Fig.1).

Step 2.2. For each DMU j_0 of the $(n_{E'} + n_F - n_{\pi'})$ DMUs with $\phi^* = 1$ and no slacks at Step 2.1, we solve a modification of problem (3.1) in which the objective is replaced by (1.3) and ϕ is fixed at $\phi = 1$ (i.e., Phase II LP). We can here identify subset E' in terms of the maximum $\sigma_{j_0}^* = 0$ and the rest of subset F in terms of $\sigma_{j_0}^* > 0$, and obtain the reference set and combination coefficients for the $(n_F - n_{\pi'})$ DMUs $j_0 \in F$.

Step 2.3. Like Step 1.3, for each DMU $j_0 \in N$ or F with $\lambda_j^* > 0$ for $j \in F$, we solve a modified form of Phase II LP of problem (3.1) in which $\phi = 1$ is replaced by $\phi = \phi^*$. Through this step and Step 2.1, we can get the reference set and combination coefficients for the $(n_N + n_{\pi'})$ DMUs $j_0 \in N$ and F.

We should here note the following: (1) θ^* for DMU $j_0 \in E$ with $\phi^* > 1$ is $\theta^* = 1$. Of course, $\theta^* = \phi^*$ for others. (2) The optimal weights u_r^*, v_i^* associated with $\theta^* = 1$ for DMU $j_0 \in E$ are calculated as u_r^* = [shadow price of constraint (3.1b)] / ϕ^*; and v_i^* = shadow price of constraint (3.1c). This is because the dual of problem (3.1) is a modification of problem (1.1) in which constraint (1.1c) for $j = j_0$ is excluded. The shadow prices obtained at Step 2.1 are the optimal weights as they are for DMU $j_0 \in E', F$ or N. (3) The step corresponding to Step 1.4 is not needed.

We can get Level 1 information by solving $(n + n_{E'} + n_F - n_{\pi'} \cong n)$ LPs throughout Steps 2.1-2.2. This means $(n_E + n_{\pi'} - n_\pi)$, i.e., nearly n_E LPs reduction from Steps 1.1-1.2. Since $(n_{E'} + n_F - n_{\pi'})$ would be very small, we may solve nearly n LPs to get Level 1 information by using the exclusion model. The number of problems to be solved at Step 2.3 may be considered equal to that at Step 1.3, so that we would solve $(n + n_{E'} + n_F + n_\alpha - n_{\pi'} \cong n)$ LPs to get Levels 2 and 3 information throughout Steps 2.1-2.3. In the DEA case mentioned in Sec.2, since $n_{E'} = n_F = 0$ (i.e., $n_\alpha = n_{\pi'} = 0$) in fact, we could get not only Level 1 but also Levels 2 and 3 information by solving 47 (= n) LPs throughout Steps 2.1-2.3.

Acknowledgments

The author would like to thank the Editor and two anonymous referees. Acknowledgments are especially due to one of the referees for his/her perceptive comments.

References

[1] D. L. Adolphson, G. C. Cornia and L. C. Walters: A unified framework for classifying DEA models. In H. E. Bradley (ed.): *Operational Research '90* (Pergamon, Oxford, 1991), 647-657.

[2] P. Andersen and N. C. Petersen: A procedure for ranking efficient units in data envelopment analysis. *Management Science*, **39** (1993) 1261-1264.

[3] A. Charnes, W. W. Cooper, A. Y. Lewin and L. M. Seiford (eds.): *Data Envelopment Analysis: Theory, Methodology, and Application* (Kluwer Academic, Boston, 1994).

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

[5] A. Charnes, W. W. Cooper and R. M. Thrall: Classifying and characterizing efficiencies and inefficiencies in data envelopment analysis. *Operations Research Letters*, **5** (1986) 105-110.

[6] A. Charnes, W. W. Cooper and R. M. Thrall: A structure for classifying and characterizing efficiency and inefficiency in data envelopment analysis. *Journal of Productivity Analysis*, **2** (1991) 197-237.

[7] A. Hashimoto: A ranked voting system using a DEA/AR exclusion model: A note. *European Journal of Operational Research*, **97** (1997) 600-604.

[8] A. Hashimoto and H. Ishikawa: Using DEA to evaluate the state of society as measured by multiple social indicators. *Socio-Economic Planning Sciences*, **27** (1993) 257-268.

[9] C. A. K. Lovell, L. C. Walters and L. L. Wood: Stratified models of education production using modified DEA and regression analysis. In A. Charnes *et al.* (eds.): *Data En-*

velopment Analysis: Theory, Methodology, and Application (Kluwer Academic, Boston, 1994), 329-351.

[10] T. Sueyoshi: Measuring the industrial performance of Chinese cities by data envelopment analysis. *Socio-Economic Planning Sciences*, **26** (1992) 75-88.

Akihiro Hashimoto
Institute of Policy and Planning Sciences
University of Tsukuba
Tsukuba, Ibaraki 305-8573, Japan
hashimot@sk.tsukuba.ac.jp