

Article

Integrated Approach to Construction of Benchmarking Network in DEA-Based Stepwise Benchmark Target Selection

Jaehun Park and Si-Il Sung *

Quality Management Operation Department, Defense Agency for Technology and Quality, 420 Dongjin-ro, Jinju-si, Gyeongsangnam-do 660-031, Korea; dudskaudts@gmail.com

* Correspondence: siil0710@dtaq.re.kr; Tel.: +82-10-3168-4876

Academic Editor: Giuseppe Ioppolo

Received: 24 April 2016; Accepted: 23 June 2016; Published: 27 June 2016

Abstract: Stepwise benchmark target selection in data envelopment analysis (DEA) is a realistic and effective method by which inefficient decision-making units (DMUs) can choose benchmarks in a stepwise manner. We propose, for the construction of a benchmarking network (i.e., a network structure consisting of an alternative sequence of benchmark targets), an approach that integrates the cross-efficiency DEA, K-means clustering and context-dependent DEA methods to minimize resource improvement pattern inconsistency in the selection of the intermediate benchmark targets (IBTs) of an inefficient DMU. The specific advantages and overall effectiveness of the proposed method were demonstrated by application to a case study of 34 actual container terminal ports and the successful determination of the stepwise benchmarking path of an inefficient DMU.

Keywords: data envelopment analysis (DEA); benchmarking network; efficiency; multi-criteria

1. Introduction

Benchmark target selection has been recognized as essential to inefficient organizations' efficiency improvements. Several studies relevant to DEA-based benchmark target selection have been conducted in various fields such as public administration [1], production and design literature [2], and business management [3]. A benchmarking process generally consists of three steps: first, identifying the company acknowledged to be the best performer; second, setting the benchmarking goals; third, implementing the best practices [4]. Identification of the best performer, which is the most important task in the benchmarking process, entails evaluation of the relative efficiencies of competitors according to multiple input and output factors. For this purpose, data envelopment analysis (DEA), a methodology for measurement of the relative efficiencies among homogeneous decision-making units (DMUs), has been utilized [5]. Specifically, DEA identifies an efficient frontier comprising Pareto-optimal DMUs along with their respective efficiency scores. DMUs on the efficient frontier can serve as empirical benchmark targets for inefficient DMUs. However, DEA has a limitation in providing benchmarking information, in that it requires an inefficient DMU to achieve its target's efficiency in a single move, which is not always feasible in practice, especially when an inefficient DMU under evaluation is far from its benchmark target DMU on the efficient frontier.

To solve this challenge in DEA, various alternative benchmarking methods have been introduced by which DMUs can select benchmark targets in a stepwise manner and, thereby, achieve progressive performance improvement. Talluri [6], based on the combination of DEA, game theory and the clustering method, presented a performance evaluation and benchmarking method to provide effective stepwise benchmarks for poorly performing business processes. He illustrated an example using the 47 real manufacturing process data and interesting insights. Alirezaee and Afsharian [7] modeled

a multi-layered efficiency evaluation model for overcoming DEA difficulties in the presence of extraordinary data, based on which a stepwise improvement approach for inefficient DMUs can be formulated. They also presented a case study using data of large Canadian bank branches which includes the types of services, sales and number of staff. Estrada et al. [8] suggested a proximity-based stepwise benchmark target selection method which utilizes a self-organizing map for clustering DMUs according to the levels of input, and a reinforcement learning algorithm is adopted to derive the optimal path. The optimal results, however, may be sensitive to the values of selected self-organizing map parameters. Suzuki and Nijkamp [9] combined the distance friction minimization models and context-dependent models. The proposed method can provide a stepwise efficiency-improving projection; however, it may give only hypothetical results (limited in practical application). Sharma and Yu [10] proposed a decision tree-based context-dependent DEA method for improving the capability and flexibility of general DEA. The same authors [11] combined a data mining and DEA approach earlier to develop a diagnostic tool for effective measurement of the efficiencies of inefficient terminals, prescribing stepwise projection (in accordance with the terminals' maximum capacity and similar input properties) as the means of reaching the frontier. The proposed model derives multiple efficient frontiers for identifying the performance factors that differentiate inefficient DMUs. Lim et al. [12] advocated, for stepwise benchmark target selection, the use of the attractiveness and progress measures of context-dependent DEA along with the consideration of feasibility. In the proposed algorithm, it searches to select only local solutions, and thereby it does not guarantee a global optimum. Park et al. [13] proposed a stepwise benchmark target selection method based on three criteria such as preference, direction and similarity. Park et al. [14] also developed a DEA-based stepwise benchmarking method that considers a minimization-improving performance measure for improving the port efficiency. Golany and Thore [15] provided a DEA model with various constraints such as cone, categorical and goal constraints, and insisted that the DEA model must be flexible. Lozano and Villa [16] and the same authors [17] provided a gradual sequence of intermediate benchmarking targets based on the iterative solving algorithm for the cases of constant return to scale and variable return to scale, respectively. Khodakarami et al. [18] proposed a gradual efficiency improvement DEA model to measure sustainability of 31 Iranian industrial parks, classified follower and pioneer industrial parks regarding their environmental performance, and directed the followers toward pioneers, gradually improving their performance with two-stage structures.

All of the above-noted methods can be considered more realistic and more effective, because these overcome the limitations of general DEA in aspects of benchmarking and propose the stepwise benchmarking DMUs for each inefficient DMU; however, they have focused primarily on the issues of how DMUs are to be stratified into multiple layers and, especially, how intermediate benchmark targets (IBTs) in leading levels for lagging-level DMUs are to be selected. What they did not consider, in respect to the selection of an inefficient DMU's IBTs, is the consistency of the resource improvement pattern. The inconsistency of the resource improvement pattern is a case in which the inefficient DMU, which wants to improve efficiency, has to reduce (increase) its inputs (outputs) to benchmark the first IBT, and then has to increase (reduce) its inputs (outputs) inversely to benchmark the second IBT (hereafter we refer to the inconsistency of the resource pattern as the zigzagging condition). Thus, the inconsistency of the resource improvement pattern in selecting IBTs can cause unnecessary and ineffective activity for gradual efficiency improvement. To address this issue, this paper proposes an integrated systematic approach that minimizes resource improvement pattern inconsistency in IBT selection for DEA-based stepwise benchmarking. The proposed approach constructs a benchmarking network, which is a network structure consisting of an alternative sequence of benchmark targets, by integrating the cross-efficiency DEA, K-means clustering and context-dependent DEA methods. The specific advantages and overall effectiveness of the proposed method were demonstrated by application to 34 container terminal ports.

This paper is organized as follows. Section 2 defines and explains resource improvement pattern inconsistency in IBT selection. Section 3 discusses the proposed method, and Section 4 details our empirical study. Finally, Section 5 summarizes our work.

2. Problem Definition

DEA models are classified with respect to the type of envelopment surface, the efficiency measurement and the orientation (input or output). There are two basic types of envelopment surfaces in DEA known as CRS (constant return-to-scale) and VRS (variable return-to-scale). Consequently, there are two DEA models: CCR (Charnes, Coopers and Rhodes), which assumes CRS, and BCC (Banker, Charnes and Cooper), which assumes VRS, depending on whether scale efficiency is incorporated or not. This paper applies the BCC model since organizations are usually operating under different returns to scale in reality. As an illustration of resource improvement pattern inconsistency in IBT selection, consider the simple numerical supermarket example introduced in Cooper et al. [19] but with additional DMUs, shown in Table 1. There are 12 DMUs in total, each consuming two inputs (x_1 : employees, x_2 : floor area) and yielding one output (y : sales). The data and relative efficiency scores from the BCC input-oriented DEA model are plotted on a two-dimensional plane in Figure 1. In addition, the relative efficiency scores from the CCR and BCC input-oriented DEA models are listed in Table A1 of Appendix A. In the supermarket example, we see that the efficiency scores from the CCR and BCC DEA models are the same because scale efficiencies of all DMUs are “1”.

Table 1. Supermarket example.

Store	A	B	C	D	E	F	G	H	I	J	K	L
x_1	2	4	8	3	4	5	5	6	6	6	6	7
x_2	4	2	1	6	3	2	6	3	5	9	4	7
y	1	1	1	1	1	1	1	1	1	1	1	1

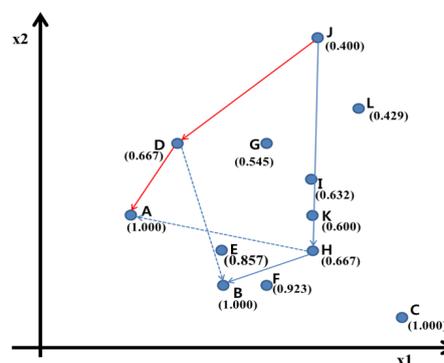


Figure 1. Sample data on a two-dimensional plane.

In Figure 1, DMUs A, B and C are the most efficient units, while the remaining nine DMUs are determined to be relatively inefficient. We suppose that relatively inefficient DMU J can choose an IBT between D and H (both have the same efficiency score: 0.667) for the stepwise benchmark target selection, and that it can then also choose an ultimate benchmark target (UBT) between A and B (both have the same efficiency score: 1). Given these assumptions, J has four alternative stepwise benchmarking paths: $J \Rightarrow D \Rightarrow A$, $J \Rightarrow D \Rightarrow B$, $J \Rightarrow H \Rightarrow A$, and $J \Rightarrow H \Rightarrow B$. Let us look at the case in which J chooses the IBTs between D and H in order to benchmark UBT A, namely $J \Rightarrow D \Rightarrow A$ and $J \Rightarrow H \Rightarrow A$. In general, an inefficient DMU that wants to improve its efficiency score (hereafter referred to as an evaluated DMU) does so by reducing its input usage or increasing its output yield under the same level of output or input in the DEA. Therefore, when J is regarded as the evaluated DMU, for the benchmarking path $J \Rightarrow D \Rightarrow A$, it has to reduce inputs (x_1, x_2) to benchmark D by 3 and 3, respectively,

and then it has to reduce inputs to benchmark A by 1 and 2, respectively. On the other hand, for the benchmarking path $J \Rightarrow H \Rightarrow A$, J has to reduce inputs (x_1, x_2) to benchmark H by 0 and 6, respectively, and then it has to reduce inputs (x_1, x_2) to benchmark A by 0 and -1 (not reduce x_2 , but inversely increase by 1), respectively. In other words, for stepwise benchmarking in the case of benchmarking path $J \Rightarrow H \Rightarrow A$, J has to reduce x_2 by 6 and then increase it by 1 inversely. In general, increasing some inputs or decreasing some outputs should not be considered an unreasonable strategy for a DMU's efficiency score improvement. However, this, particularly in the above-noted example whereby x_2 is reduced and then inversely increased, can be an unnecessary and ineffective efficiency-improvement approach. So, let us now consider the two other alternative benchmarking paths: $J \Rightarrow H \Rightarrow B$ and $J \Rightarrow D \Rightarrow B$. For the benchmarking path $J \Rightarrow H \Rightarrow B$, J has to reduce inputs (x_1, x_2) to benchmark H by 0 and 3, respectively, and then reduce inputs to benchmark B by 2 and 1, respectively. On the other hand, for the benchmarking path $J \Rightarrow D \Rightarrow B$, J has to increase input x_1 to benchmark B by 1 inversely and then reduce x_2 by 4 after it benchmarks D . In the benchmarking path $J \Rightarrow D \Rightarrow B$, D might be an inadequate IBT, since it leads to an ineffective resource-improvement activity. In other words, if J benchmarks A as its UBT, D can be a more effective and proper IBT than H in terms of resource improvement. The benchmarking paths $J \Rightarrow D \Rightarrow B$ and $J \Rightarrow H \Rightarrow A$ can be zigzagged benchmarking paths, which do not consider the consistency of the resource improvement pattern of DMU J . In other words, D , when J benchmarks B as the UBT, or H , when J benchmarks A as the UBT, are zigzagging benchmark targets. In order to minimize the selection of H when J benchmarks A as the UBT, we consider the similarity of the benchmarking direction. Consider the following case in Figure 2. In general, an inefficient DMU improves its efficiency by reducing its inputs usage or increasing its output yield under the same level of output or inputs in the DEA. We suppose that J wants to benchmark A as the UBT, and considers D and H as IBTs. The red line, indicated as L-1, represents the benchmarking path from J to A , and we can see that the benchmarking direction from J to D is closer to L-1 than that from J to H . For another case, suppose that DMU J wants to benchmark B as the UBT, and consider D and H as IBTs. We can see that the benchmarking direction from L to H is close to the benchmarking direction from J to B than that from J to D . Additionally, we can consider the resource pattern ratios. The resource $(x_1$ and $x_2)$ pattern ratios of J , A , D and H can be represented as $(1:1.5)$, $(1:2)$, $(1:2)$ and $(2:1)$, respectively. Note that the resource pattern ratios in this example can seem commensurable, but they are represented under the condition that the units of inputs 1 and 2 are 10 persons and 100 m², respectively. Here, we can see that the resource pattern ratios of A and D are similar to that of J , whereas that of H is dissimilar to that of J . Therefore, we can define that IBTs, which are located close to the benchmarking direction from the evaluated DMU to the UBT are similar in terms of the resource pattern ratio. Consequently, the similarity of the benchmarking direction to minimize the inconsistency of the resource improvement pattern in selecting IBTs can be measured by the resource pattern ratio. Although considering the similarity of the benchmarking direction cannot completely avoid the selection of inadequate DMUs as the IBT, it can reduce its probability.

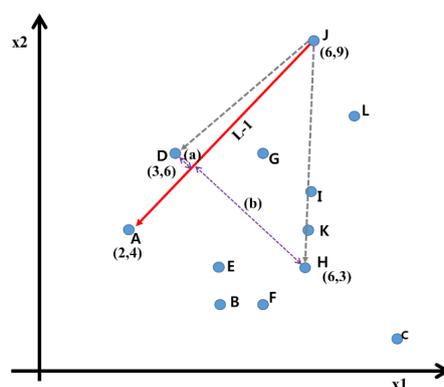


Figure 2. Similarity of benchmarking direction.

An inefficient organization generally should establish benchmarking strategies and implementation plans on how to improve efficiency after selecting its benchmarking targets. For stepwise benchmarking target selection, the inefficient organization has to establish benchmarking strategies and implementation plans at each benchmarking step. However, establishing heterogeneous benchmarking strategies and implementation plans at every benchmarking step can be a factor that decreases the benchmarking efficiencies. If an inefficient organization can select IBTs by maintaining similar benchmarking strategies and implementation plans, it might be a more effective and efficient benchmarking activity in selecting the benchmark targets gradationally. The resource improvement pattern addressed in this paper can help to maintain the benchmarking strategies and implementation plans in selecting IBTs.

3. Proposed Method

Figure 3 shows the framework of the proposed method. The proposed method is realized in sequence as follows: stratifying DMUs, clustering DMUs and constructing the benchmarking network. The method, using the stratification method in context-dependent DEA, starts by stratifying all DMUs into several layers according to their efficiency scores. Through this method, stratified benchmarking paths are obtained, and the evaluated DMU can gradually select IBTs in each layer. Then, utilizing both the cross-efficiency DEA method and the K-means clustering algorithm, all of the DMUs are classified into several clusters. Through this combined method, DMUs that are similar in terms of benchmarking direction are classified into the same cluster, and DMUs in the same cluster can be considered as the benchmarking candidate set of the evaluated DMU. Finally, a benchmarking network is constructed based on the similarity of the benchmarking direction. The benchmarking network comprises multiple stratified benchmarking paths, each of which consists of a sequence of IBTs in each layer.

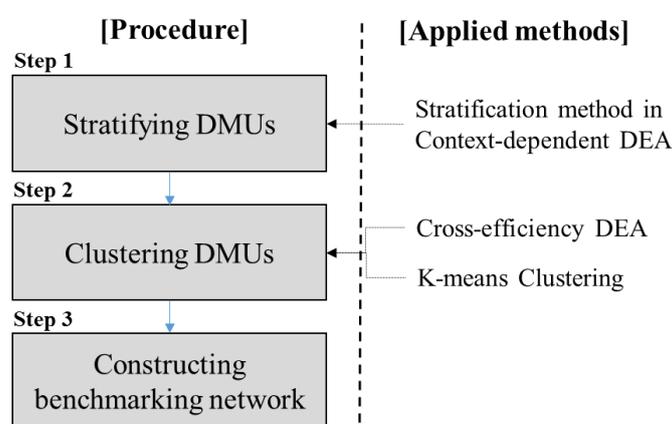


Figure 3. Framework of proposed method.

3.1. Stratification of DMUs

All of the DMUs are stratified into several layers according to their efficiency scores so that the evaluated DMU can gradually select IBTs. The stratification method in the context-dependent DEA method proposed by Seiford and Zhu [20] is utilized to stratify DMUs. Let $J^1 = \{DMU_j, j = 1, \dots, n\}$ be the set of all n DMUs, and iteratively define $J^{l+1} = J^l - E^l$, where $E^l = \{DMU_k \in J^l \mid \theta^*(l, k) = 1\}$, and $\theta^*(l, k)$ is the optimal objective value of the following linear programming model in which DMU_k is under evaluation.

$$\begin{aligned}
\theta^*(l, k) &= \min_{\lambda_j, \theta(l, k)} \theta(l, k) \\
\text{s.t.} \quad &\sum_{j \in F(J^l)} \lambda_j x_{ij} \leq \lambda(l, j) x_{ik} \\
&\sum_{j \in F(J^l)} \lambda_j y_{rj} \geq y_{rk} \\
&\sum_{j \in F(J^l)} \lambda_j = 1, \\
&j \in F(J^l)
\end{aligned} \tag{1}$$

Here, $j \in F(J^l)$ means $DMU_j \in J^l$, which is to say, $F(\cdot)$ represents the corresponding subscript index set, and E^l consists of all of the efficient DMUs on the l -th-level best-practice frontier. When $l = 1$, model (1) becomes the original input-oriented CCR (Charnes, Cooper and Rhodes) model, and the DMUs in set E^1 define the first-layer efficient frontier. Also when $l = 1$, model (1), after the exclusion of the first-layer efficient DMUs, yields the second-layer efficient frontier. In this way, the model is solved iteratively until all of the DMUs are excluded. By this process, we can identify all of the layers of the efficient frontier. The stratification method proceeds according to the following algorithm.

Step 1: Set $l = 1$, and J^l is the set of all DMUs.

Step 2: Evaluate the set of DMUs, J^l , by model (1) to obtain the l -th-layer efficient DMUs, which comprise set E^l .

Step 3: Exclude the efficient DMUs from future DEA runs: $J^{l+1} = J^l - E^l$. (If $J^{l+1} = \emptyset$, stop.)

Step 4: Evaluate the new subset of "inefficient" DMUs, J^{l+1} , by model (1) to obtain a new set of efficient DMUs E^{l+1} (the new best-practice frontier).

Step 5: Let $l = l + 1$. Go to Step 2.

Stopping rule: If $J^{l+1} = \emptyset$, the algorithm stops.

For gradual selection of benchmark targets based on the stratified layers, we specify that the evaluated DMU can sequentially select IBTs in each layer. When the above DEA stratification method is applied to the example data in Table 1, the following stratification result is obtained: E^1 (first layer) = {A, B, C}, E^2 (second layer) = {D, E, F}, E^3 (third layer) = {G, H}, E^4 (fourth layer) = {K, I}, and E^5 (fifth layer) = {J, L}.

As mentioned above, this paper applies the BCC input-oriented DEA model, which assumes VRS. Considering VRS enables an inefficient DMU to select benchmark targets that have a similar size in terms of the inputs or outputs, it consequently may help avoid inconsistent efficiency improvement strategies. Although the BCC model helps avoid inconsistent efficiency improvement, it is insufficient to minimize the resource improvement pattern inconsistency in selecting IBTs for the stepwise benchmarking target selection. Consider a benchmarking path: $J \Rightarrow E \Rightarrow A$. A is the most efficient DMU on the first layer, and E is the relative highly efficient DMU on the second layer. We identify that E can be an IBT, which causes the resource improvement pattern inconsistency for J to benchmark A because the benchmarking path $J \Rightarrow E \Rightarrow A$ can be a zigzagging path. The next sub-section proposes a method to minimize the resource improvement pattern inconsistency in selecting IBTs on each layer when an evaluated DMU benchmarks an UBT.

3.2. Classification of DMUs Based on Similarity of Benchmarking Direction

Next, we classify DMUs by performing two steps: construction of a DMU cross-efficiency matrix using the cross-efficiency DEA method proposed by Sexton et al. [21], and classification of those DMUs by the K-means clustering algorithm proposed by MacQueen [22]. Through this combined method, we define a new protocol whereby the DMUs that are similar in terms of the benchmarking direction are classified into the same cluster.

The cross-efficiency evaluation has been considered a powerful extension of DEA. Over the last several years, numerous subsequent developments have been achieved, and its usage has proliferated.

The main idea of cross-efficiency evaluation is to use DEA in a peer-evaluation mode instead of a self-evaluation mode of conventional DEA. Under cross-efficiency evaluation, mavericks have a lower chance of attaining high appraisal. Although DEA cross-efficiency evaluation has proven effective in ranking DMUs, some problems that limit its use still exist. One such well-known problem is the non-uniqueness of cross-efficiencies. Doyle and Green [23] also noted that the non-uniqueness of the DEA optimal weights generally means that the average of the peer scores is not uniquely determined, thus potentially reducing the usefulness of the methodology. Specifically, cross-efficiency scores obtained from the original DEA model are generally not unique, and depend on which of the alternate optimal solutions to the DEA linear programs is used. Several approaches have been developed to alleviate this problem by introducing secondary objectives. Doyle and Green [23] developed aggressive and benevolent model formulations. The aggressive model aims to maximize the efficiencies of all DMUs under evaluation as well as minimize the cross-efficiencies of the other DMUs. The benevolent model aims to simultaneously maximize both the efficiencies of all of the DMUs under evaluation and the cross-efficiencies of the other DMUs. More recently, Liang et al. [24] developed a game cross-efficiency to obtain a set of weights that yield cross-efficiency scores as Nash Equilibrium points, and Cook and Zhu [25] proposed to use the unit-invariant multiplicative DEA model to calculate the cross-efficiency scores that obtain maximum (and unique) cross-efficiency scores under the condition that each DMU's DEA efficiency score remains unchanged. Doyle and Green's [23] aggressive and benevolent cross-efficiency models are simple and easy to apply. Because the benevolent model has a drawback, which is its requirement that the efficiency score be raised abnormally, we apply the PEG (pairwise efficiency game) model proposed by Talluri [6], which is one of the aggressive models.

The PEG model is represented as model (2), where p is the target and θ_{pp} is the efficiency score of the p -th target DMU evaluated by model (1), which is the general DEA model. The PEG model is iteratively solved by altering the target DMU, resulting in $n - 1$ optimal weights. In the end, each DMU will have $n - 1$ optimal cross-efficiency scores given by $n - 1$ target DMUs along with its own optimal efficiency score. Thereby, each DMU constructs a cross-efficiency matrix ($J \times J$). The cross-efficiency scores of the competitor DMUs vary according to the weights of the target DMU. Applying the PEG model to the supermarket example results in the cross-efficiency matrix (12×12) shown in Table 2. More specifically, the competitor DMUs that have a similar benchmarking direction (the resource pattern ratio) will have similar cross-efficiency scores under the same target DMU. For example, the cross-efficiency scores of competitor DMUs A and D in Table 2 are very similar, because their input pattern ratios are the same (i.e., 1:2). In another case, we can identify that the cross-efficiency scores of competitor DMUs B and H, which have the same input pattern ratio, 1:0.5, also are very similar. Table 3 shows the Pearson correlation results among competition DMUs in Table 2 by the correlation analysis. In Table 3, we identify that the correlation values in $\{A, D, G, J\}$ are greater than 0.85; consequently, there is a strong positive correlation in $\{A, D, G, J\}$. In $\{A, D, G, J\}$, D and G can be IBTs of J which are located close to the benchmarking direction from J to A, and we can identify that D and G are located closer to the benchmarking direction from J to A than others in Figure 2. Similarly, there is a relatively high positive correlation in $\{B, E, F, H, I, K, L\}$, in which the correlation values are greater than 0.5. In $\{B, E, F, H, I, K, L\}$, K and F can be IBTs of L which are located close to the benchmarking direction from L to B.

Based on the cross-efficiency matrix, we classify the DMUs by applying the K-means clustering algorithm, wherein the competitor DMU represents objects to be clustered and the cross-efficiency scores are attributes describing the objects. In setting the number of k -centroids, we satisfy two conditions: the UBTs are distributed evenly in each cluster, and there are no inefficient DMUs that have any UBTs in their clusters. Because the efficient DMUs in the first layer can be regarded as the UBTs of the evaluated DMU, and since these UBTs have to be distributed evenly in each cluster and all clusters have to include at least one UBT, the number of k -centroids is equal to the number of efficient DMUs in the first layer. For example, if the number of efficient DMUs in the first layer is four, the number of k -centroids is four.

Table 2. Cross-efficiency matrix of supermarket example.

		Target DMUs											
		A	B	C	D	E	F	G	H	I	J	K	L
Competitor DMUs	A	1.000	0.500	0.250	0.667	0.500	0.400	0.400	0.333	0.286	0.333	0.333	0.286
	B	0.500	1.000	0.500	0.333	0.667	0.800	0.333	0.667	0.571	0.222	0.500	0.286
	C	0.250	0.500	1.000	0.167	0.333	0.500	0.167	0.333	0.333	0.111	0.250	0.143
	D	1.000	0.500	0.250	0.667	0.500	0.400	0.400	0.334	0.286	0.334	0.334	0.286
	E	0.643	0.857	0.429	0.429	0.857	0.686	0.429	0.571	0.490	0.286	0.571	0.367
	F	0.462	0.923	0.577	0.308	0.615	0.923	0.308	0.615	0.615	0.205	0.462	0.264
	G	0.818	0.681	0.341	0.545	0.681	0.545	0.545	0.454	0.389	0.363	0.454	0.389
	H	0.500	1.000	0.500	0.334	0.667	0.800	0.334	0.667	0.572	0.222	0.500	0.286
	I	0.474	0.948	0.553	0.316	0.632	0.885	0.316	0.632	0.632	0.211	0.474	0.271
	J	0.900	0.600	0.300	0.600	0.600	0.480	0.480	0.400	0.343	0.400	0.400	0.343
	K	0.600	0.900	0.450	0.400	0.800	0.720	0.400	0.600	0.514	0.267	0.600	0.343
	L	0.751	0.751	0.375	0.501	0.751	0.601	0.501	0.501	0.429	0.334	0.501	0.429

Table 3. Results of correlation among competitor DMUs in Table 2.

	A	B	C	D	E	F	G	H	I	J	K
B	0.058										
C	−0.245	0.473									
D	1.000	0.057	−0.246								
E	0.341	0.870	0.239	0.340							
F	−0.027	0.968	0.584	−0.027	0.781						
G	0.854	0.429	−0.140	0.854	0.725	0.312					
H	0.058	1.000	0.473	0.057	0.870	0.968	0.429				
I	−0.010	0.982	0.551	−0.010	0.806	0.997	0.338	0.982			
J	0.972	0.205	−0.234	0.972	0.508	0.100	0.946	0.205	0.122		
K	0.237	0.937	0.320	0.236	0.984	0.863	0.629	0.937	0.885	0.401	
L	0.671	0.655	0.003	0.671	0.903	0.539	0.944	0.655	0.567	0.805	0.834

$$\begin{aligned}
 & \min \sum_{r=1}^s v_r y_{rj} \\
 & \text{s.t.} \sum_{i=1}^m u_i x_{ij} = 1 \\
 & \sum_{r=1}^s v_r y_{rj} - \sum_{i=1}^m u_i x_{ij} \leq 0, \quad \forall i \neq k \\
 & \sum_{r=1}^s v_r y_{rp} - \theta_{pp} \sum_{i=1}^m u_i x_{ip} = 0 \\
 & u_i, v_r \geq 0, i = 1, \dots, m, r = 1, \dots, s, j = 1, \dots, J, p = 1, \dots, J
 \end{aligned} \tag{2}$$

Figure 4 shows the classified result with the number of *k*-centroids being three, which is the same number as the number of efficient DMUs in the first layer. *A* in the first layer is regarded as the UBT of *J*, *B* is regarded as the UBT of *L*, and the DMUs in each cluster can be considered to be similar in their benchmarking directions. We can identify that *D* and *G* in Cluster 1 are located close to the *J* to *A* benchmarking path, and that *I*, *K*, *H*, *E* and *F* in Cluster 2 are located close to the *L* to *B* benchmarking path.

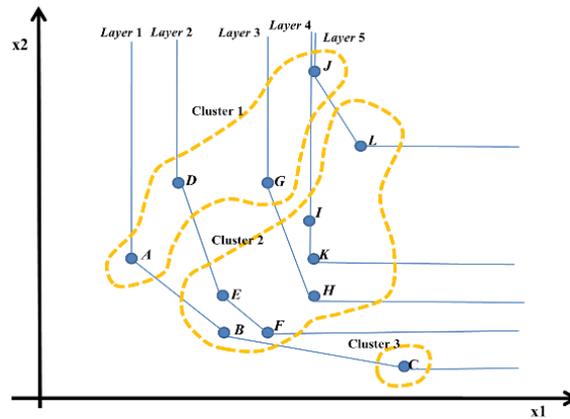


Figure 4. Classification with number of k-centroids set at three.

3.3. Benchmarking Network Construction

Based on the DMU classification, the benchmarking network is constructed. In Figure 4, A in the first layer is regarded as the UBT of J, B is regarded as the UBT of L, and the DMUs in each cluster can be considered to be similar in their resource pattern ratios. As shown in Figure 4, we can identify that D and G in Cluster 1 are located close to the J to A benchmarking path, and that I, K, H, E and F in Cluster 2 are located close to the L to B benchmarking path. Consideration of the similarity of benchmarking direction is done for the purpose of at least limiting (if not perfectly avoiding) the inconsistency of the resource improvement pattern in selecting IBTs suitable for maintaining the established gradational benchmarking strategies and implementation plans.

The benchmarking candidate set of the l -th layer is defined as $R^l = \{DMU_j \in (E^l \cap C^e) \mid l = 1, \dots, L-1\}$, where E^l is the DMU set in the l -th layer, and C^e is the DMU set in the cluster that includes the evaluated DMU. For example, in Figure 4, the benchmarking candidate sets of L are $R^1 = \{B\}$, $R^2 = \{E, F\}$, $R^3 = \{H\}$, $R^4 = \{K, I\}$, and $R^5 = \{L\}$. K or I in R^4 can be regarded as the first IBT of L in R^5 , and H in R^3 can be regarded as the second IBT of both K and I. Here, the number of benchmarking steps for an evaluated DMU depends on not only the number of layers but also on whether the benchmarking candidate set exists in the l -th layer. The benchmarking network of DMU J and L is illustrated in Figure 5. DMU J can select G and D as its IBTs and A as its UBT sequentially; DMU L can select its first IBT between K and I, H as its second IBT, its third IBT between E and F, and B as its UBT. When J benchmarks A as its UBT, the resource improvement pattern inconsistency DMUs set in layers 3 and 2 are H and E or F, respectively. In Figure 5, we can identify that H is excluded in layer 3, and E and F also are excluded in layer 2 as J's IBTs. Similarly, when L benchmarks B as its UBT, the resource improvement pattern inconsistency DMUs set in layers 3 and 2 are G and D, respectively, and we can also identify that G and D are excluded in layers 2 and 3 as L's IBTs.

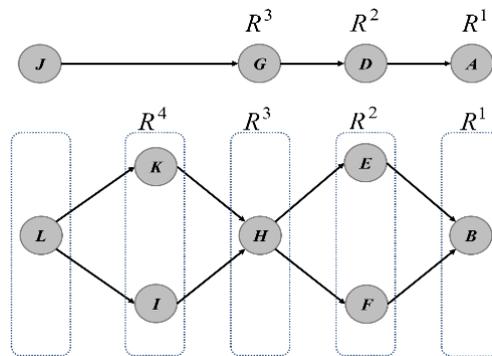


Figure 5. Benchmarking network of DMUs J and L.

4. Application

For a case study, we applied our proposed method to real container terminal port data in order to select the stepwise benchmarking path of an inefficient port under evaluation. The relevant data sources had been collected for 34 container terminal ports from the *Containerization International Year Book 2005–2007* [26], excluding incomplete data. We programmed the DEA and cross-efficiency DEA models to assess the relative efficiencies of DMUs and construct the DMU cross-efficiency matrix using software Lingo 10 [27]. To verify the programed result, we compared the efficiency scores with the DEAFrontier tool [28].

As specified in Vis and de Koster [29], when a vessel arrives at the port, quay cranes (QCs) take the import containers off the ship's hold or off the deck. Next, the containers are transferred from the QCs to vehicles such as cranes or straddle carriers (SCs) that ravel between the ship and the stack. A straddle carrier can both transport containers and store them in the stack. If a vehicle arrives at the stack, it puts the load down or the stack crane takes the container off the vehicle and stores it in the stack. After a certain period, the containers are retrieved from the stack by cranes and transported by vehicles to transportation modes such as barges, deep sea ships, trucks or trains. To load export containers onto a ship, these processes are executed in reverse order. The general container-handling process can be categorized as four working stages: dispatching, loading, receiving (gate-in), delivering (gate-out). Each working stage has its own workflow, which consists of several tasks, and this container operation process is nearly similar irrespective of terminals or ports worldwide.

As mentioned in Sharma and Yu [10], the terminal area and quay length are the best input variables for the "land" factor, and the number of quay gantry cranes, the number of yard gantry cranes, the number of straddle carriers and the number of reach stackers are the best input variables for the "equipment" factor in the container terminal operation. The container throughput is the most appropriate and analytically tractable output variable for the container terminal efficiency evaluation. In this paper, the berth length (m), total area (m³), CFS (container freight station) and the number of loading machines were used as inputs, while the number of unloading TEU (Twenty-foot equivalent unit) and the number of loading TEU were applied as outputs for the efficiency evaluation of the DMUs. Container gantries, yard gantries, a quay crane, a floating crane, a mobile crane, a straddle carrier, a forklift, a reach stacker, a top lifter, a yard tractor and a yard trailer were considered as loading machines, while both empty and full containers were included in the number of unloading and loading containers. The ports were regarded as DMUs. The descriptive statistics on the input and output data and the 34 ports' relative efficiency scores attained through classical DEA are listed in Tables 4 and 5, respectively.

Table 4. Descriptive statistics for inputs and outputs used.

	Resource	Min.	Max.	Avg.	Std. Dev.
Inputs	Berth length	181	15,585	4054	3210
	Total area	105,000	7,156,000	1,697,593	1,524,231
	CFS area	5000	1,535,000	161,850	288,312
	Loading machine	8	1955	363	369
Outputs	Unloading TEU	46,393	7,180,397	1,280,869	1,639,527
	Loading TEU	44,537	7,386,658	1,290,066	1,650,742

Table 5. Relative efficiency scores for 34 ports.

Port No.	Port Name	θ	Reference set	Port No.	Port Name	θ	Reference set
1	Hong Kong	1.000	-	18	Colombo	0.924	1, 2, 26
2	Singapore	1.000	-	19	Bangkok	0.303	1, 26
3	Kaohsiung	1.000	-	20	Osaka	0.281	1, 2, 3, 26
4	Rotterdam	0.297	1, 2, 26	21	Hampton Roads	0.339	1, 2, 3
5	Busan	0.818	1, 31	22	Charleston	0.281	1, 3
6	Hamburg	0.260	1, 3	23	Melbourne	0.217	1, 2, 3
7	Yokohama	0.369	1, 2, 3, 26	24	La Spezia	0.389	1, 2, 26
8	Los Angeles	0.492	2, 26	25	Genoa	0.310	1, 2, 3
9	Antwerp	0.152	1, 2, 3	26	Laem Chabang	1.000	-
10	New York	0.147	1, 2	27	Qingdao	0.670	2, 3, 26
11	Dubai	0.421	1, 2	28	Southampton	0.441	1, 2, 3
12	Keelung	0.319	1, 2, 26	29	Santos	0.752	1, 31
13	Manila	0.183	1, 2, 26	30	Barcelona	0.207	1, 2, 3, 26
14	Oakland	0.213	1, 2, 3	31	Jeddah	1.000	-
15	Seattle	0.184	1, 2, 26	32	Sydney	0.055	1, 2, 3, 26
16	Tanjung Priok	0.559	1, 2	33	Khor Fakkan	0.510	2, 26
17	Port Klang	0.233	1, 2, 26	34	Valencia	0.149	1, 2

In the relative efficiency assessment, five ports (Hong Kong, Singapore, Kaohsiung, Laem Chabang and Jeddah) were determined to be the most efficient (efficiency score = 1), while the remaining 29 ports were determined to be inefficient. Six ports (Antwerp, New York, Manila, Seattle, Sydney and Valencia) were determined to be the most inefficient (efficiency score < 0.2). The reference set is a set of the most efficient ports, and it is regarded as the benchmarking targets of inefficient ports for efficient improvement. For instance, let us take the inefficient port Valencia (efficiency score of 0.149); this particular terminal is referenced to Hong Kong or Singapore. Although Valencia can select Hong Kong or Singapore as the benchmarking target, it would be difficult to benchmark it in a single step due to the high efficiency gap. From the port stratification, six layers of the efficient frontier were determined. In the first layer, the five most efficient ports (Hong Kong, Singapore, Kaohsiung, Laem Chabang and Jeddah) were included. We classified the ports with k-centroids into five classes, which is the same number as the ports in the first layer. The stratification and classification of all of the ports is schematized in Figure 6. We could determine that the ports in the first layer were evenly distributed among the clusters. We selected Valencia as the evaluated DMU; its UBT was Singapore. The benchmarking candidate sets were determined to be $R^4 = \{\text{Seattle, Rotterdam, Oakland}\}$, $R^3 = \{\text{Dubai, Southampton, La Spezia}\}$, $R^2 = \{\text{Colombo, Qingdao, Khor Fakkan}\}$, and $R^1 = \{\text{Singapore}\}$. Valencia's benchmarking network was constructed as shown in Figure 7. In a parallel case, New York, Antwerp and Keelung's benchmarking networks were constructed as shown in Figures 8–10, respectively.

As mentioned in Section 2, considering the resource improvement pattern can help to maintain the benchmarking strategies and implementation plans for an evaluated DMU to benchmark an UBT, and it can be a more effective and efficient benchmarking activity for the stepwise benchmarking target selection. In the case of New York, Melbourne or Hamburg can be New York's first IBT. Then, New York has to establish benchmarking strategies and plans to benchmark one of them. Sequentially, New York can benchmark its second IBT between Bangkok and Santos, then benchmark Busan as its third IBT and finally benchmark Hong Kong as its UBT by maintaining similar benchmarking strategies and plans established in the first IBT selection. In other words, New York can implement effective and efficient benchmarking activity since it benchmarks similar stepwise benchmark targets in terms of the resource pattern similarity.

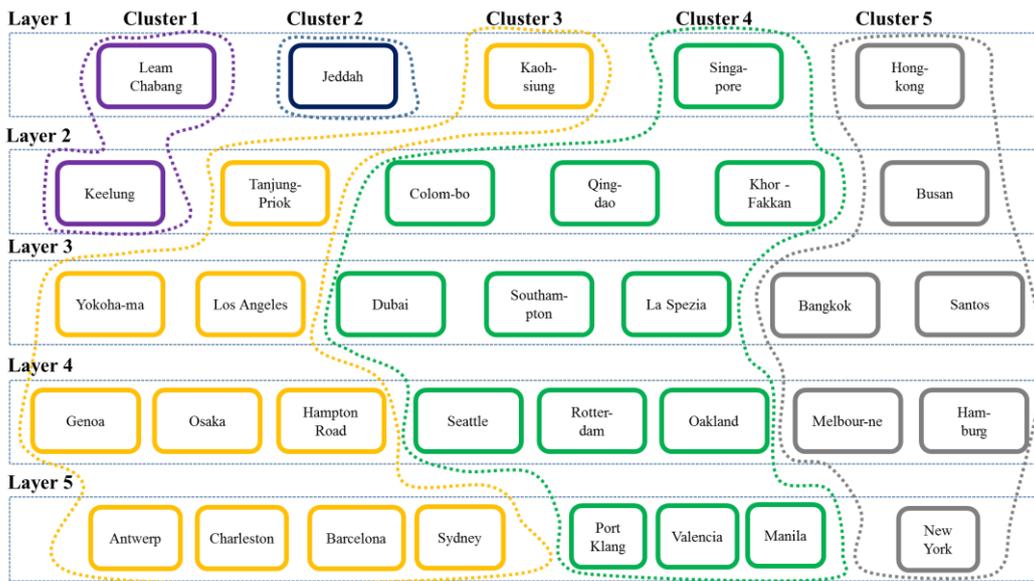


Figure 6. Stratification and clustering results for 34 container terminal ports.

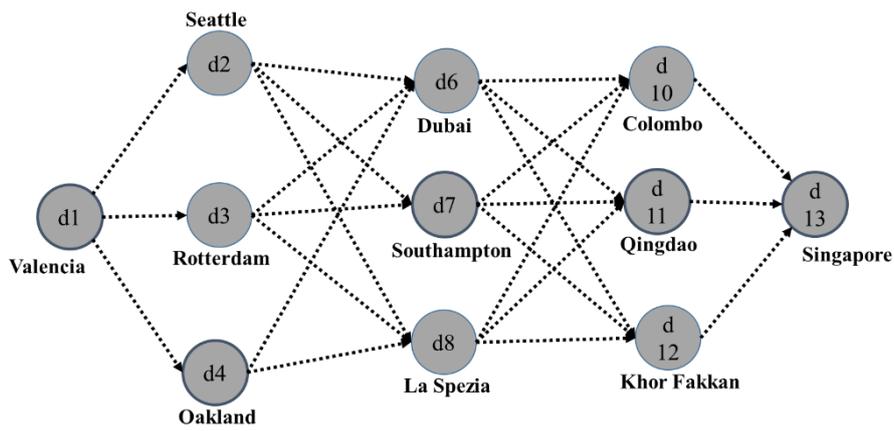


Figure 7. Benchmarking network of Valencia.

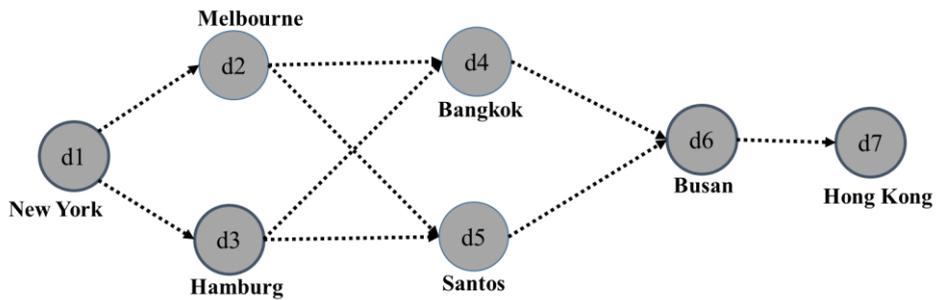


Figure 8. Benchmarking network of New York.

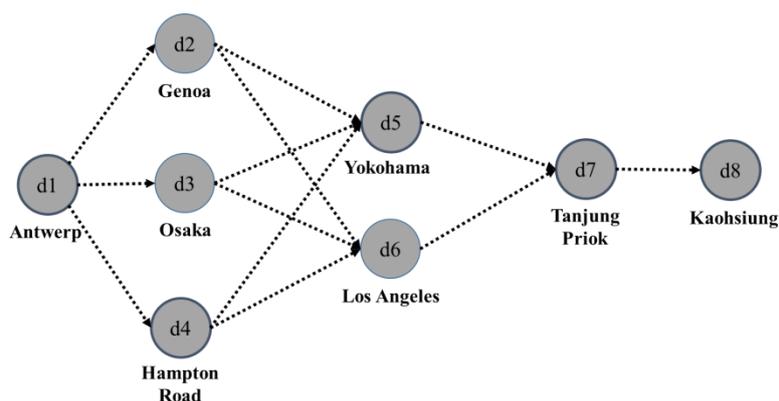


Figure 9. Benchmarking network of Antwerp.

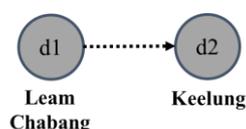


Figure 10. Benchmarking network of Leam Chabang.

5. Concluding Remarks

Stepwise benchmarking target selection in DEA is a realistic and effective method by which inefficient DMUs can choose benchmarks in a stepwise manner. However, previous DEA-based stepwise benchmarking target selection methods have focused primarily on the issues of how DMUs are to be stratified into multiple layers and, especially, on how intermediate benchmark targets (IBTs) in leading levels for lagging-level DMUs are to be selected. Those methods did not consider, with respect to the selection of an inefficient DMU's IBTs, the consistency of the resource improvement pattern. In this paper, we proposed an integrated systematic approach to the construction of a benchmarking network, which is a network structure consisting of an alternative sequence of benchmark targets. The proposed approach integrates the cross-efficiency DEA, *K*-means clustering and context-dependent DEA methods in considering the consistency of the resource improvement pattern and selecting, on that basis, the IBTs of an inefficient DMU. The proposed method was realized in sequence as follows: stratifying DMUs, clustering DMUs and constructing the benchmarking network by combining the cross-efficiency DEA and *K*-means clustering method. In this paper, we defined a new protocol whereby the DMUs that are similar in terms of the benchmarking direction are classified into the same cluster. As an application of the proposed method, a benchmarking network for the 34 container terminal ports was tested. The berth length (m), total area (m³), CFS (container freight station) and number of loading machines were used as inputs, while the number of unloading TEU and the number of loading TEU were applied as outputs for the efficiency evaluation of the DMUs, and the ports were regarded as DMUs. We constructed benchmarking networks for the inefficient ports (Valencia, New York, Antwerp and Leam Chabang) to benchmark their UBTs (Singapore, Hong Kong, Kaohsiung and Keelung), respectively. We argued that considering the resource improvement pattern can help to maintain the benchmarking strategies and implementation plans for an evaluated DMU to benchmark a UBT, and it can be a more effective and efficient benchmarking activity for the stepwise benchmarking target selection. We expect that, based on the proposed method, an effective benchmark target selection process can be established.

In spite of the utility of the proposed methodology, it does not suggest which DMU in each layer can be the best benchmark target for the evaluated DMU to reach the UBT. Neither does it consider the minimal number of stepwise benchmark targets for the inefficient DMU to reach the UBT. However, in an actual inefficient organization, the number of benchmarking steps can be an important

decision factor for stepwise efficiency improvement. If there are too many stepwise benchmark targets, a significant practical difficulty for the DMU in accomplishing the benchmarking schedule could be incurred. Therefore, how to choose the best IBT among the number of DMUs in each layer in an order so that the evaluated DMU can reach the UBT, and how to apply the number of stepwise benchmark targets for more practical stepwise benchmarking, will be issues for future research.

Author Contributions: Jaehun Park and Si-Il Sung conceived of and designed the research. Jaehun Park analyzed the data and wrote the paper. Si-Il Sung verified the results and responded to the reviewers. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DEA	data envelopment analysis
DMUs	decision-making units
IBTs	intermediate benchmark targets
UBT	ultimate benchmark target

Appendix A

Table A1. Relative efficiency scores of the supermarket example by CCR and BCC models.

Store	x_1	x_2	y	Efficiency			
				CRS	VRS	SE	Returns to Scale
A	2	4	1	1.00	1.00	1.00	CRS
B	4	2	1	1.00	1.00	1.00	CRS
C	8	1	1	1.00	1.00	1.00	CRS
D	3	6	1	0.66	0.66	1.00	CRS
E	4	3	1	0.85	0.85	1.00	CRS
F	5	2	1	0.92	0.92	1.00	CRS
G	5	6	1	0.54	0.54	1.00	CRS
H	6	3	1	0.66	0.66	1.00	CRS
I	6	5	1	0.54	0.54	1.00	CRS
J	6	9	1	0.40	0.40	1.00	CRS
K	6	4	1	0.60	0.60	1.00	CRS
L	7	7	1	0.42	0.42	1.00	CRS

DEAFrontier [28] software was used to assess the efficiency scores.

References

- Ammons, D.N. Benchmarking as a performance management tool: Experiences among municipalities in North Carolina. *J. Public Budg. Account. Financ. Manag.* **2000**, *12*, 106–124.
- Grupp, H. Technometrics as a missing link in science and technology indicators. In *Measuring the Dynamics of Technological Change*; Pinter Pub Ltd: London, UK, 1990; pp. 57–76.
- Tata, J.; Prasad, S.; Motwani, J. Benchmarking quality management practices: US versus Costa Rica. *Multinatl. Bus. Rev.* **2000**, *8*, 37.
- Donthu, N.; Hershberger, E.K.; Osmonbekov, T. Benchmarking marketing productivity using data envelopment analysis. *J. Bus. Res.* **2005**, *58*, 1474–1482. [[CrossRef](#)]
- Ross, A.; Droge, C. An integrated benchmarking approach to distribution center performance using DEA modeling. *J. Op. Manag.* **2002**, *20*, 19–32. [[CrossRef](#)]
- Talluri, S. A benchmarking method for business-process reengineering and improvement. *Int. J. Flex. Manuf. Syst.* **2000**, *12*, 291–304. [[CrossRef](#)]
- Alirezadee, M.; Afsharian, M. Model improvement for computational difficulties of DEA technique in the presence of special DMUs. *Appl. Math. Comput.* **2007**, *186*, 1600–1611. [[CrossRef](#)]

8. Estrada, S.A.; Song, H.S.; Kim, Y.A.; Namn, S.H.; Kang, S.C. A method of stepwise benchmarking for inefficient DMUs based on the proximity-based target selection. *Expert Syst. Appl.* **2009**, *36*, 11595–11604. [[CrossRef](#)]
9. Suzuki, S.; Nijkamp, P. A stepwise-projection data envelopment analysis for public transport operations in Japan. *Let. Spat. Resour.Sci.* **2011**, *4*, 139–156. [[CrossRef](#)]
10. Sharma, M.J.; Yu, S.J. Benchmark optimization and attribute identification for improvement of container terminals. *Eur. J. Op. Res.* **2010**, *201*, 568–580. [[CrossRef](#)]
11. Sharma, M.J.; Yu, S.J. Performance based stratification and clustering for benchmarking of container terminals. *Expert Syst. Appl.* **2009**, *36*, 5016–5022. [[CrossRef](#)]
12. Lim, S.; Bae, H.; Lee, L.H. A study on the selection of benchmarking paths in DEA. *Expert Syst. Appl.* **2011**, *38*, 7665–7673. [[CrossRef](#)]
13. Park, J.; Bae, H.; Lim, S. A DEA-based method of stepwise benchmark target selection with preference, direction and similarity criteria. *Int. J. Innov. Comput. Inf. Control* **2012**, *8*, 5821–5834.
14. Park, J.; Lim, S.; Bae, H. DEA-based port efficiency improvement and stepwise benchmarking target selection. *Inf. Int. Interdiscip. J.* **2012**, *15*, 6155–6172.
15. Golany, B.; Thore, S. Restricted best practice selection in DEA: An overview with a case study evaluating the socio-economic performance of nations. *Ann. Op. Res.* **1997**, *73*, 117–140. [[CrossRef](#)]
16. Lozano, S.; Villa, G. Determining a sequence of targets in DEA. *J. Op. Res. Soc.* **2005**, *56*, 1439–1447. [[CrossRef](#)]
17. Lozano, S.; Villa, G. Gradual technical and scale efficiency improvement in DEA. *Ann. Op. Res.* **2010**, *173*, 123–136. [[CrossRef](#)]
18. Khodakarami, M.; Shabani, A.; Saen, R.F. A new look at measuring sustainability of industrial parks: A two-stage data envelopment analysis approach. *Clean Technol. Environ. Policy* **2014**, *16*, 1577–1596. [[CrossRef](#)]
19. Cooper, W.W.; Seiford, L.M.; Tone, K. *Introduction to Data Envelopment Analysis and Its Uses: With DEA-Solver Software and References*; Springer Science & Business Media: Berlin, Germany, 2006.
20. Seiford, L.M.; Zhu, J. Context-dependent data envelopment analysis—Measuring attractiveness and progress. *Omega* **2003**, *31*, 397–408. [[CrossRef](#)]
21. Sexton, T.R.; Silkman, R.H.; Hogan, A.J. Data envelopment analysis: Critique and extensions. *New Dir. Progr. Eval.* **1986**, *1986*, 73–105. [[CrossRef](#)]
22. MacQueen, J. *Some Methods for Classification and Analysis of Multivariate Observations*; University of California Press: Berkeley, CA, USA, 1967; pp. 281–297.
23. Doyle, J.; Green, R. Efficiency and Cross-Efficiency in DEA: Derivations, Meanings and Uses. *J. Op. Res. Soc.* **1994**, *45*, 567–578. [[CrossRef](#)]
24. Liang, L.; Wu, J.; Cook, W.D.; Zhu, J. The DEA game cross-efficiency model and its Nash equilibrium. *Op. Res.* **2008**, *56*, 1278–1288. [[CrossRef](#)]
25. Cook, W.; Zhu, J. DEA Cobb-Douglas frontier and cross-efficiency. *J. Op. Res. Soc.* **2013**, *65*, 265–268. [[CrossRef](#)]
26. Containerization International. Available online: <https://www.lloydlist.com> (accessed on 24 June 2016).
27. Microsoft Lingo 10. Available online: <http://lingo.software.informer.com> (accessed on 24 June 2016).
28. DEAFrontier. Available online: <http://www.deafrontier.net> (accessed on 24 June 2016).
29. Vis, I.F.A.; de Koster, R. Transshipment of containers at a container terminal: An overview. *Eur. J. Op. Res.* **2003**, *147*, 1–16. [[CrossRef](#)]

