OPERATIONS RESEARCH
Vol. 61, No. 3, May–June 2013, pp. 666–676
ISSN 0030-364X (print) | ISSN 1526-5463 (online)
http://dx.doi.org/10.1287/opre.2013.1173

Data Envelopment Analysis with Nonhomogeneous DMUs

Wade D. Cook
Schulich School of Business, York University, Toronto, Ontario M3J 1P3, Canada, wcook@schulich.yorku.ca

Julie Harrison
Department of Accounting and Finance, University of Auckland, Auckland 1142, New Zealand, j.harrison@auckland.ac.nz

Raha Imanirad
Schulich School of Business, York University, Toronto, Ontario M3J 1P3, Canada, rimanirad09@schulich.yorku.ca

Paul Rouse
Department of Accounting and Finance, University of Auckland, Auckland 1142, New Zealand, p.rouse@auckland.ac.nz

Joe Zhu
School of Business, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, jzhu@wpi.edu

Data envelopment analysis (DEA), as originally proposed, is a methodology for evaluating the relative efficiencies of a set of *homogeneous* decision-making units (DMUs) in the sense that each uses the same input and output measures (in varying amounts from one DMU to another). In some situations, however, the assumption of homogeneity among DMUs may not apply. As an example, consider the case where the DMUs are plants in the same industry that may not all produce the same products. Evaluating efficiencies in the absence of homogeneity gives rise to the issue of how to fairly compare a DMU to other units, some of which may not be exactly in the same "business." A related problem, and one that has been examined extensively in the literature, is the *missing data* problem; a DMU produces a certain output, but its value is not known. One approach taken to address this problem is to "create" a value for the missing output (e.g., substituting zero, or by taking the average of known values), and use it to fill in the gaps. In the present setting, however, the issue isn't that the data for the output is missing for certain DMUs, but rather that the output isn't produced. We argue herein that if a DMU has chosen not to produce a certain output, or for any reason cannot produce that output, and therefore does not put the resources in place to do so, then it would be inappropriate to artificially assign that DMU a zero value or some "average" value for the nonexistent factor. Specifically, the desire is to fairly evaluate a DMU for what it does, rather than penalize or credit it for what it doesn't do. In the current paper we present DEA-based models for evaluating the relative efficiencies of a set of DMUs where the requirement of homogeneity is relaxed. We then use these models to examine the efficiencies of a set of manufacturing plants.

Subject classifications: nonhomogeneous DMUs; missing outputs; subgroups; assurance regions.
Area of review: Decision Analysis.
History: Received November 2011; revision received May 2012; accepted February 2013.

1. Introduction

Data envelopment analysis (DEA), as originally proposed by Charnes et al. (1978), is a methodology for evaluating the relative efficiencies of a set of *homogeneous* decision-making units (DMUs) belonging to the same technology in the sense that each uses the same inputs and outputs, measured the same way (in varying amounts from one DMU to another). In some situations, however, the assumption of homogeneity among DMUs may not apply, even though they use the same technology. As an example, consider the case where the DMUs are plants in the same industry that may not all produce the same products. Another is a set of universities, where not all have the same departments; hence, they are not homogeneous. In the current paper we present DEA-based models for evaluating the relative efficiencies of a set of DMUs that belong to the same technology, but where the requirement of homogeneity is relaxed. We then use these models to examine the efficiencies of a set of manufacturing plants.

Evaluating efficiencies in the absence of homogeneity gives rise to the issue of how to fairly compare a DMU to other units, some of which may not be exactly in the same "business." A related problem, and one that has been examined extensively in the literature, is the *missing data* problem; a DMU produces a certain output, but its value is not known. One approach taken to address this problem is to "create" a value for the missing output (e.g., by taking the average of known values), and use it to fill in the gaps. For outputs, using zero as a dummy for blank entries is another prescribed solution. The question of blank output entries is thus closely related to the treatment of zeros in the data matrices (see e.g., Thompson et al. 1993 for discussion).

In the present setting, however, the issue isn't that the data for the output is missing for certain DMUs, but rather that the output isn't produced. In the case of the universities acting as the DMUs, those without engineering departments cannot be directly compared to those that do have such departments, and substituting a value such as zero for this "missing" data is not appropriate. We argue herein that if a DMU has chosen not to produce a certain output (e.g., the missing engineering department), or for any reason cannot produce that output, and therefore does not put the resources in place to do so, then it would be inappropriate to artificially assign that DMU a zero value or some "average" value for the nonexistent factor. Specifically, the desire is to fairly evaluate a DMU for what it does, rather than penalize or credit it for what it doesn't do.

Potentially, the nonhomogeneous DMU issue could be handled by breaking the set of DMUs into multiple groups, with all members of any group producing the same outputs, and then doing a separate DEA analysis for each group. In this way, a DMU is evaluated against only *true peers*, specifically those whose output profiles are identical to its own. No attempt would be made to compare a DMU to other "partial peers," namely, those whose output profiles overlap with, but are not identical to, those of the said DMU. There are at least two problems with this approach. One is a small sample issue in that there may be, in some cases, very few (if any) actual peers. Specifically, in some situations this would require the set of DMUs to be split into multiple small sets to reflect the permutations. The greater the number of splits required, the more difficult it is to estimate meaningful efficiency. It would commonly mean that efficiency scores would be artificially inflated. Another problem is that true best practices for a DMU may in fact be those practices adopted by the partial peers, and excluding consideration of the latter may result in a failure to identify such best practices. This being the case, we wish, wherever possible, to include all DMUs in the comparison set.

Section 2 describes a problem setting involving the evaluation of a set of manufacturing plants, where identifiable groups of DMUs produce only proper subsets of the full set of outputs. Section 3 is devoted to the development of a DEA-type model for handling the general missing output situation. Generally, this is brought about by viewing the DMU as consisting of mutually exclusive subgroups of outputs. One important extension of the DEA concept that has been discussed extensively in the literature is that involving the imposition of multiplier restrictions, in particular those based upon assurance regions (AR). In a setting where there is lack of homogeneity among DMUs, such AR constraints can be problematic in that multiple and often inconsistent sets of restrictions may materialize out of the abovementioned output subgroups. Section 4 extends the new DEA methodology to allow for consideration of such conflicting AR constraints. Section 5 looks into other issues that may arise relating to nonhomogeneous DMUs, and suggests ways of handling such issues. Section 6 applies the new methodology to data for a set of 47 plants relating to the steel fabrication industry, as discussed above. Conclusions appear in §7. Supplemental material to this paper is available at http://dx.doi.org/10.1287/opre.2013.1173.

2. Manufacturing Plants with Variable Output Sets

To demonstrate the problem of nonhomogeneity of DMUs in DEA, a set of 47 steel fabrication plants is considered. The main product lines manufactured by the plants consist of the following:

1. Sheet steel products (ladders, guards, bumpers, and conveyors);
2. Flat bar products used mainly in building construction (brackets, base plates, headers, and posts);
3. Pipes and cylinders (storm drains, plumbing products, etc.);
4. Furnace and air conditioning ducts;
5. Structural steel (e.g., joists and support beams);
6. Tanks (residential and industrial).

In addition, resources employed by all plants are comprised of: (1) plant labor; (2) shears and saws; (3) presses and rolling equipment; and (4) cutting torches and welding equipment.

In this particular industry some plants choose not to manufacture certain products. As shown in Figure 1, plants with similar product lines have been grouped together into P DMU groups N_p $p = 1, \dots, P$, where in our particular case $P = 4$. Observe, for example, that plants in N_1 manufacture products 1, 2, 3, 5; those in N_2 make products 2, 3, 4, 5, 6;

Figure 1. Product lines by DMU group.

	Outputs					
Group	Sheet steel (1)	Flat bar (2)	Pipes/ cylinders (3)	Cylindrical bearings (4)	Structural steel (5)	Storage tanks (6)
N_1	X	X	X		X	
N_2		X	X	X	X	X
N_3			X		X	X
N_4	X		X		X	

etc. Part of the reason for the variability of products across a business (DMU) has to do with the focus on industrial versus residential clientele. Some companies also may cater more to sectors such as automotive than is true of others.

In the following section we develop a DEA-based methodology for dealing with nonhomogeneous settings such as that represented by Figure 1.

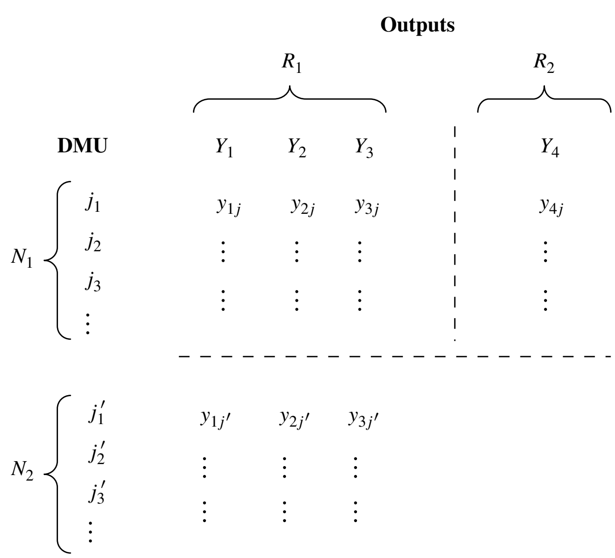
3. A DEA Model for DMUs with Variable Output Sets

In an earlier paper (Cook et al. 2012), a simple case where DMUs appeared in a 2-group setting was explored, and affords a convenient and transparent backdrop and introduction for demonstrating the methodology to be developed herein. For completeness, we summarize some of the elements of that earlier development. Specifically, consider the situation where n DMUs are organized into two subgroups N_1 and N_2, with those in N_1 producing four outputs y_1, y_2, y_3, y_4, whereas those in N_2 produce only three outputs y_1, y_2, y_3, with both subgroups using the same inputs. Figure 2 demonstrates the split of DMUs across subgroups N_1 and N_2.

Hence, when we want to evaluate a DMU, say in the first group N_1, we argue that the evaluation may reasonably be undertaken by carrying out a separate DEA analysis on each of that DMUs 2 output subgroups $R_1 = \{y_1, y_2, y_3\}$, and $R_2 = \{y_4\}$. We further argue that for DMUs in N_1, one may think of each input as being split between the production of the subset of outputs in R_1 and those in R_2. (The situation where some inputs are not separable is discussed later in §6). If we knew what the proportional split of inputs was between these two output groups, we could proceed in three stages as follows:

Stage 1: In this stage we decide on a split of the inputs across the output subgroups. For the moment and to facilitate transparency, let us assume it is known that $\alpha_{1N_1} = 90\%$

Figure 2. A two-group setting.



of the inputs for any DMU in N_1 go toward the production of R_1, and that the remaining $\alpha_{2N_1} = 10\%$ go toward the production of outputs in R_2. We generalize this idea below.

Stage 2: In this stage we derive, for each DMU, efficiency scores for the individual subgroups making up that DMU. Specifically, take 90% of the inputs held by each DMU in N_1 and carry out a standard DEA analysis of *all n* DMUs (using outputs in R_1). Here, whenever we are looking at a DMU j in N_1, we need to remember we have replaced the original amounts of its inputs x_{ij} by the proportional amounts of these $\tilde{x}^1_{ij} = \alpha_{1N_1} x_{ij}$ (that have been assigned to the outputs in R_1). Note, as well, that in this simple case, the inputs held by DMUs in N_2 do not need to be split up, because there is only one relevant subgroup of outputs (R_1). Hence, in this case, $\alpha_{1N_2} = 1$ and $\alpha_{2N_2} = 0$. Carry out a standard DEA analysis of each of the *members* in N_1 using the outputs in R_2 and with inputs $\tilde{x}^2_{ij} = \alpha_{2N_1} x_{ij}$. Recall that we do not include the members from N_2 in this analysis, because these DMUs do not produce outputs in R_2.

Stage 3: For DMUs in N_2, the DEA scores arrived at in stage 2 are the final scores. For DMUs in N_1 we combine the scores from steps 1 and 2 by taking a weighted average (discussed below).

We point out that this idea of splitting inputs across various subsets of outputs is similar in nature to the methodology developed for uncovering multiple variable proportionality (MVP), as described in Cook and Zhu (2011).

It is reasonable to argue at this point that rather than following the above procedure, one might instead simply assign to DMUs in N_2 a value of zero for the missing output y_4 and proceed with a conventional DEA analysis. Perhaps the best counterargument to this is that the DMUs in N_2 are, under a conventional DEA analysis, at liberty to assign a zero weight to those outputs (y_4 in this case) that are at a zero level, thereby inferring in a mathematical sense that the DMUs in N_1 are in the same "business" as those in N_2. The problem with this is that DMUs in N_2, with their limited product line, would commonly use fewer resources than is the case for their *full-service* peers in N_1 ... less labor, less machine time, less inventory carrying cost, etc. Hence, DMUs in N_2 are accorded an unfair advantage over their N_1 peers. To illustrate, consider the simple example where two DMUs have the following profiles:

DMU no.	y_1	y_2	x
1	100	100	20
2	100		20

Here, DMU 1 is producing 100 units of each of two products, utilizing 20 units of a single input, whereas DMU 2 uses the same amount of input, but produces only 100 units of output 1. Clearly, under a conventional DEA analysis, both DMUs will be deemed efficient given that DMU 2 can assign a zero multiplier to the second output. Suppose, however, that we knew that approximately 80% of

DMU 1's input went toward the production of product 1, and the remaining 20% was used to produce product 2. Let us present the above data in a more exact manner, replacing DMU 1 by two sub-DMUs that we call DMU 1(a) and DMU 1(b).

DMU no.	y_1	y_2	x
1(a)	100		16
1(b)		100	4
2	100		20

Now, following the above notation, R_1 is the output set consisting of y_1, and R_2 contains the output y_2. DMU 2 is now evaluated properly against DMU 1(a), and since the latter uses less input than the former, the input-oriented efficiency score for DMU 2 is only 0.8.

Hence, our argument is that in evaluating the efficiency of DMU 2 (or in general those in N_2), the comparison to DMU 1 (or those in N_1), should be against only that part of DMU 1's business that it has in common with DMU 2.

The General Case

Let us now examine the general setting, and as a working example, consider the situation portrayed in Figure 1, where a set of manufacturing plants produces a certain set of products, but not all products are produced in all plants. Suppose the plants fall into P mutually exclusive (M.E.) groups, as described in §2, which we denote by $\{N_p\}_{p=1}^{P}$. Here $P=4$.

Now form M.E. output subgroups R_k, $k=1,\ldots K$, where R_k denotes the subset of outputs with the property that all of its members appear as the outputs of exactly the same set of DMUs (same DMU "profile"). Specifically, if outputs $r_1, r_2 \in R_k$, then the DMU profiles of these two outputs are identical. Hence, if r_1 is an output for DMU groups 1, 2, 4, then r_2 is an output for exactly the same DMU groups. Also, each R_k is *maximal* in the sense that there is no output $r \notin R_k$ that has the same DMU profile as members of R_k. It can be shown that for the above DMU profiles, the K output sets are

$$R_1=\{1\},\quad R_2=\{2\},\quad R_3=\{3,5\},\quad R_4=\{4\},\quad R_5=\{6\}.$$

A general algorithm for deriving the maximal output groupings is found in Appendix A.

Theorem 3.1. *The generated set of maximal output subgroups is unique.*

Proof. Let us assume that the set of maximal output subgroups is not unique. In that case there must exist at least two different sets of output subgroups S_1 and S_2. It can then be implied that there must be at least one R_k in S_1 that is different from R_k in S_2. Consequently, there must exist at least one output $r \in R_k$ in S_1 such that $r \notin R_k$ in S_2. This proves that R_k in S_2 is not maximal because there exists output $r \notin R_k$ that has the same DMU profile as members of R_k. Hence, it can be concluded that for each k there exists only one maximal R_k, and as a result there can only be one set of maximal output subgroups. This completes the proof.

Definition 3.1. Let L_{N_p} denote those R_k forming the full output set for any DMU in N_p.

In the steel plant setting, $L_{N_1}=\{R_1, R_2, R_3\}$, $L_{N_2}=\{R_2, R_3, R_4, R_5\}$, $L_{N_3}=\{R_3, R_5\}$, $L_{N_4}=\{R_1, R_3\}$.

To evaluate the efficiency of a given DMU, we need to proceed in three stages. In stage 1 we decide (for the DMU under evaluation, say $j_o \in N_{p^o}$) what portion of each input i will be allocated to each of the output subgroups $R_k \in L_{N_{p^o}}$; we denote this proportion by $\alpha_{iR_kp^o}$. In stage 2 we evaluate the efficiency of the DMU in terms of each of its subgroups R_k, and in stage 3 we take a weighted average of these subgroup scores to get the overall efficiency of the DMU.

Stage 1: *Deriving the Split of Inputs*. Let us formalize the ideas for the situation where we *do not* know the precise split of resources as was assumed above. Let the decision variable α_{iR_kp} denote the proportion of input i to be allocated to outputs in subgroup R_k of L_{N_p}. We argue that the best way to divide up the resources, hence determining the most appropriate alpha variables, is to do so in a manner that results in the best overall or aggregate score for the DMU, across all of its business subunits. Further, we argue that the overall efficiency of a DMU $j_o \in N_{p^o}$ can reasonably be represented as a weighted average (convex combination) of the R_k-subgroup efficiencies (across all output subgroups in N_{p^o}). We point out that this argument is essentially that the DMU is the sum of its parts, and therefore assumes there are no economies or diseconomies of scope. In cases where it is believed such economies (diseconomies) of scope exist, our approach may not accurately capture efficiency at the aggregate level.

Given that it is aggregate efficiency of the DMU that we wish to derive, and that this aggregate will be represented as a convex combination of the R_k-subgroup efficiencies, we set out to determine the α-split of inputs with the objective of maximizing this aggregate efficiency. With this in mind, consider the following input-oriented radial projection model for a DMU $j_o \in N_{p^o}$. It is noted that the development in this section is, in the spirit of Charnes et al. (1978), presented from the perspective of the constant returns to scale (CRS) technology. As demonstrated in a later section, however, the concepts are equally valid in a variable returns to scale (VRS) setting.

$$e_o = \max \sum_{R_k \in L_{N_{p^o}}} W_{R_k j_o}\left[\sum_{r\in R_k} u_r y_{rj_o} \Big/ \sum_i \nu_i \alpha_{iR_kp^o} x_{ij_o}\right], \quad (1a)$$

$$\text{subject to} \sum_{R_k \in L_{N_p}} W_{R_k j}\left[\sum_{r\in R_k} u_r y_{rj} \Big/ \sum_i \nu_i \alpha_{iR_kp} x_{ij}\right] \leqslant 1$$

$$\forall\, j \in N_p,\ R_k \in L_{N_p},\ p=1,\ldots,P, \quad (1b)$$

$$\sum_{r\in R_k} u_r y_{rj} - \sum_i \nu_i \alpha_{iR_kp} x_{ij} \leqslant 0 \quad \forall\, j\in N_p,\ R_k\in L_{N_p},\ p=1,\ldots,P, \quad (1c)$$

$$\sum_{R_k\in L_{N_p}} \alpha_{iR_kp} = 1 \quad \forall\, i,\ p=1,\ldots,P, \quad (1d)$$

$$a_{iR_kp} \leqslant \alpha_{iR_kp} \leqslant b_{iR_kp} \quad \forall\, i, R_k, p=1,\ldots,P, \quad (1e)$$

$$u_r, \nu_i, \alpha_{iR_kp} \geqslant 0, \quad \forall\, i,\ R_k,\ p. \quad (1f)$$

We point out that whereas in the above example it was assumed that the same values of alpha applied to all DMUs, in the general case here, the model makes provision for a different set of alpha variables for each DMU j. The basic idea of this model is to represent the overall efficiency of a DMU as a convex combination ($\sum_{R_k\in L_{N_{p^o}}} W_{R_kj_o}=1$) of the efficiencies $\sum_{r\in R_k} u_r y_{rj_o} / \sum_i \nu_i \alpha_{iR_kp^o} x_{ij_o}$ of the individual subgroups R_k. Although the weights $W_{R_kj_o}$ may be any set of values that represent the importance to be attached to the relevant subgroups, there would appear to be at least two reasonable and obvious choices. From an accounting perspective, it is appropriate and reasonable to let the proportion of inputs assigned to (or consumed by) a subgroup dictate the importance of that subgroup to the overall DMU; the subgroup assigned the largest share of resources would be given the highest weight. An equally valid definition of importance of a subgroup would be to base it upon the proportion of the aggregate output for the DMU generated by that subgroup; the subgroup that creates the greatest value for the DMU would be weighted the highest. One might also adopt a net contribution or profit criterion to select weights. As a convenience in the case of the input-oriented model adopted herein, we select the first of these two approaches, namely, we base the weights for the subgroup ratios on the proportions of the aggregate inputs consumed by those subgroups. Thus, we define the weight $W_{R_kj_o}$ to be assigned to subgroup R_k as

$$W_{R_kj_o} = \sum_i \nu_i \alpha_{iR_kp^o} x_{ij_o} \Big/ \sum_{R_k\in L_{N_{p^o}}} \left[\sum_i \nu_i \alpha_{iR_kp^o} x_{ij_o}\right]. \quad (2)$$

Constraints (1b) require that the multipliers chosen for a DMU j_o satisfy the condition that when they are applied to any other DMU, the corresponding ratio (of outputs to inputs) does not exceed unity. At the same time, and in anticipation of the second stage, we impose the requirement that the ratio of outputs to inputs at the *subgroup level* also not exceed unity. Specifically, constraints (1c) specify that the resource-splitting variables α_{iR_kp} be selected in a manner that allows the efficiency ratio corresponding to the subset of outputs in R_k to assume a value that does not exceed unity for some values of the multipliers u_r, ν_i. We note that in the presence of (1c), constraints (1b) are redundant and may be dropped from the model.

Constraints (1d) specify that the α values assigned to the subgroups of outputs corresponding to any set p sum to unity for each i. Finally, constraints (1e) place lower and upper limits on the sizes of the α variables. It is worth noting that in a situation wherein a particular input may not in fact impact certain outputs or output subgroups, the corresponding α_{iR_kp} can of course be set to zero.

The Equivalent Linear Formulation

Problem (1) in its current form is nonlinear. To facilitate linearization, first note that by virtue of the definition we choose to use for the $W_{R_kj_o}$ as given by (2), the objective function (1a) becomes

$$e_o = \max\left[\sum_{R_k\in L_{N_{p^o}}} \sum_{r\in R_k} u_r y_{rj_o} \Big/ \sum_i \nu_i x_{ij_o}\right]. \quad (1a')$$

Specifically, maximizing the weighted average of subgroup ratios is equivalent to maximizing the overall efficiency ratio of the DMU.

Now make the change of variables $z_{iR_kp} = \nu_i \alpha_{iR_kp}$, and note that

$$\sum_{R_k\in L_{N_p}} \alpha_{iR_kp} = 1 \ \Rightarrow\ \nu_i \sum_{R_k\in L_{N_p}} \alpha_{iR_kp} = \nu_i \ \Rightarrow\ \sum_{R_k\in L_{N_p}} z_{iR_kp} = \nu_i$$

Using the usual transformation $t = 1/\sum_i \nu_i x_{ij_o}$ (see Charnes et al. 1978), and defining $\mu_r = tu_r$, $\upsilon_i = t\nu_i$, $\gamma_{iR_kp} = tz_{iR_kp}$, problem (1) becomes

$$e_o = \max \sum_{R_k\in L_{N_{p^o}}} \sum_{r\in R_k} \mu_r y_{rj_o} \quad (3a)$$

subject to
$$\sum_{R_k\in L_{N_{p^o}}} \left(\sum_i \gamma_{iR_kp^o} x_{ij^o}\right) = 1, \quad (3b)$$

$$\sum_{r\in R_k} \mu_r y_{rj} - \sum_i \gamma_{iR_kp} x_{ij} \leqslant 0 \quad \forall\, j\in N_p,\ R_k\in L_{N_p},\ p=1,\ldots,P, \quad (3c)$$

$$\sum_{R_k\in L_{N_p}} \gamma_{iR_kp} = \upsilon_i \quad \forall\, i,\ p=1,\ldots,P, \quad (3d)$$

$$\upsilon_i a_{iR_kp} \leqslant \gamma_{iR_kp} \leqslant \upsilon_i b_{iR_kp} \quad \forall\, i,\ R_k\in L_{N_p},\ p=1,\ldots,P, \quad (3e)$$

$$\mu_r, \upsilon_i, \gamma_{iR_kp} \geqslant \varepsilon, \quad \forall\, r, i, R_k, p=1,\ldots,P. \quad (3f)$$

Stage 2: *Deriving the Subgroup Efficiency Scores*. Note that the purpose of stage 1 is to derive, for each DMU j_o in N_{p^o}, the "optimal" proportions of inputs $\hat{\alpha}_{iR_kp^o}$ to be assigned to output subgroups R_k. These are given by $\hat{\alpha}_{iR_kp^o} = \hat{\gamma}_{iR_kp^o}/\hat{\upsilon}_i$. When these proportions are available (from the solution to Model (3)), one can then allocate to subgroup R_k the appropriate amount of input x_{ij_o}, namely $\tilde{x}^k_{ij_o} = \hat{\alpha}_{iR_kp^o} x_{ij_o}$. The conventional CCR DEA model (see Charnes et al. 1978) can then be applied to each of the subgroups R_k of j_o. Specifically, determine M_{R_k}, the set of all DMU groups that have R_k as a member, that is

$$M_{R_k} = \{N_p \text{ such that } R_k \in L_{N_p}\}. \quad (4)$$

Note, for example, in the six-output steel fabrication application described above, $M_{R_1} = \{N_1, N_4\}$, $M_{R_2} = \{N_1, N_2\}, \ldots,$ etc.

Now, for each DMU j_o, and each subgroup R_{k° corresponding to the set N_{p^o} that contains j_o as a member, solve the DEA model:

$$\begin{aligned} e_{R_{k^\circ} j_o} = \max \quad & \sum_{r \in R_{k^\circ}} \mu_r y_{rj_o} \\ \text{subject to} \quad & \sum_i \upsilon_i \tilde{x}_{ij_o}^{k^\circ} = 1, \\ & \sum_{r \in R_{k^\circ}} \mu_r y_{rj} - \sum_i \upsilon_i \tilde{x}_{ij}^{k^\circ} \leqslant 0 \\ & \qquad j \in N_p, \text{ for } N_p \in M_{R_{k^\circ}}, \\ & \mu_r, \upsilon_i \geqslant \varepsilon. \end{aligned} \tag{5}$$

Stage 3: *Deriving the Aggregate Efficiencies*. The overall efficiency score of the DMU j_o is now derived by taking a weighted average of the subgroup scores obtained in stage 2, using the $W_{R_k j_o}$ defined in (2). It should be pointed out that in computing $W_{R_k j_o}$ an appropriate set of input multipliers ν_i needs to be chosen. Furthermore, the multipliers need to be computed in an environment where all subunits are being compared simultaneously. The aggregate Model (3) provides such an environment. That is, in (3) when DMU j_o is being evaluated, the input portion of expression (3c), namely $\sum_i \gamma_{iR_k p^o} x_{ij_o}$ (for $j = j_o$), represents the *value* of that DMU's resources that are assigned to subgroup R_k. The *total value* of all resources consumed by DMU j_o is given by $\sum_i \upsilon_{ij_o} x_{ij_o}$, which is scaled to unity as per constraint (3b). Hence, the weights $W_{R_k j_o}$ reduce to $W_{R_k j_o} = \sum_i \gamma_{iR_k p^o} x_{ij_o}$. Note again that this set of weights is dependent on the particular DMU j_o under investigation, to reflect the fact that the proportion of inputs allocated to the kth subunit is DMU specific.

The model developed in this section permits one to evaluate efficiencies of a set of DMUs where output profiles are not homogeneous across those units. The proposed approach portrays a DMU's performance as a convex combination of its component parts (subgroups). It is important to point out that in the above structure we do not consider restrictions that might be imposed on the multipliers μ, υ (referring to Model (3)), other than those that restrict efficiency ratios to not exceed unity. This is raised here because such restrictions may lead to infeasibilities that would normally not occur in the conventional DEA setting. The following section investigates the role that multiplier restrictions play in this more general environment. First, however, we point to related literature.

Relation to Previous Work

The methodology developed above is related to two strands of previous research. First, network DEA, as originated by Färe and Grosskopf (1996), sets out to evaluate DMU performance by examining the internal subprocesses that make up the DMU. Whereas one can define performance in many ways, if one concentrates on technical efficiency, network DEA provides for both subprocess efficiency scores as well as an overall score for the DMU itself. Thus, the approach herein is a form of network DEA analysis in that the subprocesses are the subunits as we have described above. Arguably, one difference between our methodology and that characterizing network DEA is that our definition of the overall performance of the DMU is that it is a weighted average of the subunit efficiencies. What is normally done in network DEA is to use a conventional DEA model to describe overall efficiency in terms of all inputs entering the DMU versus all outputs leaving the DMU. As well, subprocess shares of inputs would normally be known in advance (except in allocative efficiency settings), as opposed to those shares being derived as part of the optimization procedure, as is the situation herein. Furthermore, there is no clear direct connection in network DEA between the efficiency score for the overall DMU and the scores of the subprocesses. We provide that connection in the methodology presented here.

Other related research carried out by Cook and Hababou (2001) and by Cook et al. (2000) is closer still to work done herein. In that former work, the DMUs are bank branches that are viewed as consisting of two components or subunits, namely sales and service. Those authors develop an overall efficiency score for the branch using a model analogous to (1) above. Their model sets out to optimize the ratio of total weighted outputs to total weighted inputs for the overall branch. Component efficiencies (sales and service) are then simply taken as the ratio of weighted outputs to inputs for those components, $\sum_{r \in R_c} \mu_r y_{rj} / \sum_{i \in I} \gamma_{ik} x_{ij}$ similar to expression (3c). Here R_c denotes the output bundles for either the sales or service component. The problem with using this ratio to capture component efficiency is that it doesn't properly capture the component's performance. Specifically, since it is overall branch performance that is being maximized, there is no internal mechanism for insuring that at the same time component scores are appropriately set, consistent with DEA constructs. Our model herein takes the next important step (step 2 above) of using the resource split across the subunits to find the maximal efficiencies for each of those subunits, and then taking the weighted average of those maximal scores (step 3) to arrive at the score for the DMU. In addition, our methodology herein identifies the subunits into which to decompose the DMU, whereas the earlier research pertained only to those applications where components or subunits are well defined in advance. Finally, the earlier work did not consider the issue of conflicting AR constraints as we do herein. This topic is covered in the next section.

4. AR Restrictions on Pairs of Input Variables

Many different forms of multiplier restrictions in DEA analyses have been discussed in the literature, but none more than those that take the form of assurance regions (AR). AR constraints, as first discussed by Thompson et al. (1990), involve the placing of bounds on the ratios of pairs of multipliers. The resulting DEA-AR model has been employed extensively in numerous performance measurement settings. In this section, we address the problem of nonhomogenous DMUs in the presence of such AR restrictions on input multipliers, and the inherent problems that can arise thereon. It should be noted that although the discussion focuses on input multipliers, the concepts apply equally to the output side.

Let us assume, in reference to Model (3), that for each output subgroup R_k AR constraints of the form $c_{iL}^k v_{i_2} \leqslant v_{i_1} \leqslant c_{iU}^k v_{i_2}, k = 1, \ldots, K$, have been specified. As indicated above, such constraints identify the relative magnitudes of pairs of input multipliers v_{i_1} to v_{i_2} within output subgroup R_k. It can be argued that all such constraints across all output subgroups R_k can be expressed in the form $c_{iL}^k \leqslant v_i/v_n \leqslant c_{iU}^k$, such that v_n is the designated *numeraire*, the multiplier for one of the inputs against which all other multipliers are compared (see Thompson et al. 1990). Given that subgroups of outputs are in some senses a signal that multiple business units are operating under one umbrella, it is often the case that multiple sets of AR constraints on any given pair of multipliers can emerge simultaneously. Moreover, such multiple sets can result in infeasibility. For example, let us assume that the following constraints $3 \leqslant v_2/v_1 \leqslant 5$ and $6 \leqslant v_2/v_1 \leqslant 8$ have been specified for output subgroups R_{k_1} and R_{k_2}, respectively. If these two sets of restrictions were to be imposed simultaneously on Model (3), infeasibility would obviously result. This being the case, there is reason to look for a mechanism that will permit one to fold such multiple sets of constraints involving any multiplier v_i into a single set, thereby insuring that Model (3) is feasible.

Assume that AR constraints $c_{iL}^{k_1} \leqslant v_i/v_n \leqslant c_{iU}^{k_1}$ and $c_{iL}^{k_2} \leqslant v_i/v_n \leqslant c_{iU}^{k_2}$ have been imposed within R_{k_1} and R_{k_2}, respectively. To reduce this pair of AR restrictions to a single expression, we propose focusing attention on one of the bounds, say the lower bound. It is observed that by expressing

$$c_{iL}^{k_2} \leqslant v_i/v_n \leqslant c_{iU}^{k_2} \tag{6}$$

in the form $(c_{iL}^{k_1}/c_{iL}^{k_2})c_{iL}^{k_2} \leqslant (c_{iL}^{k_1}/c_{iL}^{k_2})v_i/v_n \leqslant (c_{iL}^{k_1}/c_{iL}^{k_2})c_{iU}^{k_2}$, and by subsequently making the following transformation $v_i' = (c_{iL}^{k_1}/c_{iL}^{k_2})v_i$, (6) can be converted to $c_{iL}^{k_1} \leqslant v_i'/v_n \leqslant (c_{iL}^{k_1}/c_{iL}^{k_2})c_{iU}^{k_2}$. Consequently, for each DMU, $j_o \in M_{R_{k_2}}$, $v_i x_{ij}$ can be replaced by $(c_{iL}^{k_1}/c_{iL}^{k_2})v_i x_{ij}(c_{iL}^{k_2}/c_{iL}^{k_1})$ or $v_i' x_{ij}(c_{iL}^{k_2}/c_{iL}^{k_1})$, meaning that by scaling the multiplier v_i by a factor $c_{iL}^{k_1}/c_{iL}^{k_2}$, we can scale the data for x_i in $M_{R_{k_2}}$ by the reciprocal of that factor.

To illustrate, refer again to the above example of constraints $3 \leqslant v_2/v_1 \leqslant 5$ and $6 \leqslant v_2/v_1 \leqslant 8$ in subgroups R_{k_1} and R_{k_2}, respectively, and assume that input 1 is used as the numeraire. To arrive at a single set of constraints involving v_2, we first replace the constraint $6 \leqslant v_2/v_1 \leqslant 8$ with $\frac{3}{6}6 \leqslant \frac{3}{6}v_2/v_1 \leqslant \frac{3}{6}8$. By making the transformation $v_2' = \frac{3}{6}v_2$, we can then replace $v_2 x_2$ in $M_{R_{k_2}}$ with $\frac{3}{6}v_2(\frac{6}{3}x_2)$ or $v_2'(\frac{6}{3}x_2)$.

Specifically, by scaling v_2 down by a factor $\frac{3}{6}$ we can scale up the data for x_2 in subgroup $M_{R_{k_2}}$ by a factor $\frac{6}{3}$. Along the same lines, the upper bound on v_i/v_n is replaced by $\bar{c}_{iU}^{k_2} = (c_{iL}^{k_1}/c_{iL}^{k_2})c_{iU}^{k_2}$.

This exercise is then repeated for all other output subgroups that have AR constraints involving multipliers v_i and v_n. Let us define

$$\bar{c}_{iL} = c_{iL}^{k_1} \tag{7}$$

$$\bar{c}_{iU} = \min\{\bar{c}_{iU}^{k_1}, \bar{c}_{iU}^{k_2}, \ldots\}, \tag{8}$$

where it is understood that the minimum in (8) is taken over all R_k that contain an AR constraint involving the two variables v_i and v_n. Expression (6) can now be replaced by

$$\bar{c}_{iL} \leqslant \frac{v_i}{v_n} \leqslant \bar{c}_{iU}. \tag{9}$$

To repeat, assume a set of AR restrictions on a pair of input variables (v_i, v_n) has been imposed within various output subgroups R_k. That is, the AR restrictions can vary by output subgroup. Let one of these subgroups, $R_{\hat{k}}$, be the base against which all other sets will be compared. As a result of the adjustments made to reduce these multiple restrictions to a single AR constraint, the corresponding inverse adjustments must be made to variable x_i within each of the R_k subgroups. (We are assuming that v_n is the designated numeraire for this pair of variables). Let us now denote the adjusted input data by x_{ikj}, that is,

$$x_{ikj} = (c_{iL}^k/c_{iL}^{\hat{k}})x_{ij}. \tag{10}$$

With these adjustments having been made to the input data, Model (3) for a given DMU j_o in DMU group N_{p^o} now takes the form

$$e_o = \max \sum_{R_k \in L_{N_{p^o}}} \sum_{r \in R_k} \mu_r y_{rj_o} \tag{11a}$$

$$\text{subject to} \sum_{R_k \in L_{N_{p^o}}} \Big(\sum_i \gamma_{iR_k p^o} x_{ikj^o}\Big) = 1, \tag{11b}$$

$$\sum_{r \in R_k} \mu_r y_{rj} - \sum_i \gamma_{iR_k p} x_{ikj} \leqslant 0 \quad \forall j \in N_p,\ R_k \in L_{N_p},\ p = 1, \ldots, P, \tag{11c}$$

$$\sum_{R_k \in L_{N_p}} \gamma_{iR_k p} = v_i \quad \forall i,\ p = 1, \ldots, P, \tag{11d}$$

$$v_i a_{iR_k p} \leqslant \gamma_{iR_k p} \leqslant v_i b_{iR_k p} \quad \forall i,\ R_k \in L_{N_p},\ p = 1, \ldots, P, \tag{11e}$$

$$v_n \bar{c}_{iL} \leqslant v_i \leqslant v_n \bar{c}_{iU} \quad \forall i,\ i \neq n, \tag{11f}$$

$$\mu_r, v_i, \gamma_{iR_k p} \geqslant \varepsilon, \quad \forall r, i, R_k, p = 1, \ldots, P. \tag{11g}$$

Note that (11f) represents the final constraints resulting from the amalgamation of the multiple AR restrictions corresponding to the various R_k subgroups.

In the case of the stage 2 subgroup optimization, where the efficiency of subgroup R_{k^o} is to be determined, the AR-equivalent of Model (5) is given by (12). Here, it is important to note that only AR restrictions relevant to this particular output subgroup, and no AR restrictions outside this subgroup, are invoked. This being the case, input data requires AR-adjustment only in cases where multiple AR restrictions on a pair of variables arise. This latter can happen when constraints (on a given pair of variables) are invoked in one of the output subgroups R_k that are different from those invoked in another subgroup. We use the notation $\hat{x}_{ij}^{k^o}$ to denote the alpha-adjusted, and AR-adjusted version of input x_i consumed by output subgroup R_{k^o}. Constraints (12d) reflect the imposed AR constraints. The following model is now solved for each of the output subgroups, R_{k^o} for each DMU j_o:

$$e_{R_{k^o} j_o} = \max \sum_{r \in R_{k^o}} \mu_r y_{rj_o} \tag{12a}$$

$$\text{subject to} \quad \sum_i \upsilon_i \hat{x}_{ij_o}^{k^o} = 1, \tag{12b}$$

$$\sum_{r \in R_{k^o}} \mu_r y_{rj} - \sum_i \upsilon_i \hat{x}_{ij}^{k^o} \leqslant 0 \quad j \in N_p, \text{ for } N_p \in M_{R_{k^\circ}}, \tag{12c}$$

$$\upsilon_n c_{iL}^{k^\circ} \leqslant \upsilon_i \leqslant \upsilon_n c_{iU}^{k^\circ} \quad \forall i,\ i \neq n, \tag{12d}$$

$$\mu_r, \upsilon_i \geqslant \varepsilon.$$

5. Other Considerations

5.1. Nonseparable Inputs

In many instances there can be inputs that do not lend themselves to subdivision in the manner described above. If, for example, in the analysis of the steel fabrication plants, one wished to include as an input a quality measure pertaining to supplier reliability, it would appear to be unreasonable to suggest subdividing this factor, and assigning portions of it across the various subunits; such a factor, in its entirety, would affect the outputs in each subunit k. Generalizing, let us use the notation I_s, I_{ns} to denote the sets of separable and nonseparable inputs, respectively. In the discussion thus far, all inputs have been assumed to belong to I_s. The efficiency ratio for a given subgroup R_k within DMU j_o can now be expressed in the form $\sum_{r \in R_k} u_r y_{rj_o} / (\sum_{i \in I_s} \nu_i \alpha_{iR_k} x_{ij_o} + \sum_{i \in I_{ns}} \nu_i^{R_k} x_{ij_o})$, where $\nu_i^{R_k}$ is the worth or weight assigned to the nonseparable input x_{ij_o}, $i \in I_{ns}$, and represents the impact of that input on the outputs in R_k. Note that we are permitting this weight to be different from one subgroup to another. Following the logic of (2), we define the weight attached to the efficiency ratio for subgroup R_k by

$$W_{R_k j_o} = \left[\sum_{i \in I_s} \nu_i \alpha_{iR_k p^o} x_{ij_o} + \sum_{i \in I_{ns}} \nu_i^{R_k} x_{ij_o} \right] \Big/ \sum_{R_k \in L_{N_{p^o}}} \left[\sum_{i \in I_s} \nu_i \alpha_{iR_k p^o} x_{ij_o} + \sum_{i \in I_{ns}} \nu_i^{R_k} x_{ij_o} \right]. \tag{13}$$

The optimization model for this more general case would be identical in form to (3) with the exception that constraints (3b) and (3c) are replaced by

$$\sum_{i \in I_s} \upsilon_i x_{ij_o} + \sum_{R_k \in N_{p^o}} \left(\sum_{i \in I_{ns}} \upsilon_i^{R_k} \right) x_{ij_o} = 1 \tag{4b$'$}$$

and

$$\sum_{r \in R_k} \mu_r y_{rj} - \sum_{i \in I_s} \gamma_{ki} x_{ij} - \sum_{i \in I_{ns}} \upsilon_i^{R_k} x_{ij} \leqslant 0, \quad \forall j \in N_p, \tag{3c$'$}$$

respectively, to account for the two types of inputs. Furthermore, constraints (3d) and (3e) apply only to separable inputs i. Here, $\upsilon_i^{R_k} = t \nu_i^{R_k}$ under the usual transformation as discussed above.

5.2. Variable Returns to Scale

The development above is based on a CRS technology. In the situation where a VRS technology is deemed to be more appropriate, it is sufficient to replace terms such as $\sum_{r \in R_k} u_r y_{rj}$ by $\sum_{r \in R_k} u_r y_{rj} - u^o$. An advantage of the VRS formulation is that the sign of u^o is subgroup dependent, and signals whether the projected version of that (sub)DMU will be experiencing increasing, constant, or decreasing returns to scale. This can provide useful information to management regarding the returns to scale orientation of various parts of the business, and may aid in deciding how to redistribute resources, given that common resources (I_s) are shared among the subgroups.

5.3. Output Orientation

The development throughout has assumed that efficiency is to be viewed from the perspective of an input orientation. If the organization intends to improve efficiency by pursuing output expansion rather than input reduction, then an output orientation would be an appropriate model structure to use. Specifically, Model (1) would be replaced by

$$e_o = \min \sum_{R_k \in L_{N_{p^o}}} W_{R_k j_o} \left[\sum_i \nu_i \alpha_{iR_k p^o} x_{ij_o} \Big/ \sum_{r \in R_k} u_r y_{rj_o} \right] \tag{13a}$$

$$\text{subject to} \sum_{R_k \in L_{N_p}} W_{R_k j} \left[\sum_i \nu_i \alpha_{iR_k p} x_{ij} \Big/ \sum_{r \in R_k} u_r y_{rj} \right] \geqslant 1 \quad \forall j \in N_p,\ R_k \in L_{N_p},\ p = 1, \ldots, P, \tag{13b}$$

$$\sum_i \nu_i \alpha_{iR_k p} x_{ij} - \sum_{r \in R_k} u_r y_{rj} \geqslant 0 \quad \forall j \in N_p,\ R_k \in L_{N_p},\ p = 1, \ldots, P, \tag{13c}$$

$$\sum_{R_k \in L_{N_p}} \alpha_{iR_k p} = 1 \quad \forall i,\ p = 1 \ldots, P, \tag{13d}$$

$$a_{iR_k p} \leqslant \alpha_{iR_k p} \leqslant b_{iR_k p} \quad \forall i,\ R_k,\ p = 1, \ldots, P, \tag{13e}$$

$$u_r, \nu_i, \alpha_{iR_k p} \geqslant 0, \quad \forall i,\ R_k,\ p. \tag{13f}$$

As discussed earlier, it appears equally valid to base the definition of weights $W_{R_k j_o}$ on either inputs consumed or outputs generated. In the case of the output-oriented model, it is therefore reasonable to weight the efficiency ratios according to the latter (outputs generated). Specifically, if the weight on subgroup R_k is chosen as

$$W_{R_k j_o} = \sum_{r \in R_k} u_r y_{rj_o} \Big/ \sum_{R_k \in L_{N_{p^o}}} \left[\sum_{r \in R_k} u_r y_{rj_o} \right], \tag{14}$$

then the above aggregate objective function (13a) becomes the ratio of overall DMU input to overall DMU output, namely,

$$e_o = \min \sum_{R_k \in L_{N_p o}} \left[\sum_i \nu_i x_{ij_o} \Big/ \sum_{r \in R_k} u_r y_{rj_o} \right]. \quad (13a')$$

We now apply the above methodology to the derivation of efficiencies of a set of steel fabrication plants.

6. Application

To demonstrate the application of the models developed in the earlier sections, data on a set of 47 plants, as per Appendix C, Tables C.1 and C.2, were considered. These plants are grouped into four DMU subgroups N_1 to N_4 such that plants belonging to any DMU group N_p produce identical products. The profiles of the four groups of DMUs N_p are as described in §2. For example, DMUs in N_1 produce outputs 1, 2, 3, and 5. Note from Appendix C, Table C.1, that N_1 consists of DMUs 1, 3, 4, 6, 9, 10, 11, 19, 32, 35, 39, and 46.

In the application considered herein, AR restrictions play an important role, particularly on the input side. They provide a way to bring resource trade-offs into the picture. Input multipliers effectively mimic resource costs. Hence, whereas it is the case in many real-world settings that the development of such restrictions can be problematic, in manufacturing situations resource costs often provide the appropriate route to deriving the desired restrictions. To that end, data were collected relating to per-unit costs for each of the four inputs. The data provided in Table 1 represent per-unit costs incurred during the last quarter of 2010. For example, the range of cost estimates specified for labor (x_1) for the last quarter of 2010 is from \$5,000 to \$7,500 per plant employee (wages and benefits). Although there is no implied variation in labor costs across the five bundles, $k = 1, 2, 3, 4, 5$, wage rates can differ from plant to plant and over time due to the mix of full-time and part-time labor used. For this reason a range is given for this input.

For the other three inputs, machine “rates” were assumed to be the estimated quarterly costs of depreciation, routine maintenance, and unforeseen breakdown costs. In the case of the shearing machines, for example, the estimated quarterly cost (depreciation and maintenance) of operating one machine would generally vary between \$7,000 and \$10,000 per quarter in the case of output bundle $k = 1$, and \$14,000 and \$19,000 in the case of $k = 2$. The increased stress placed on the equipment in the production of flat bar products versus that created in the manufacture of sheet steel products, contributes to the difference in cost between the two product groupings.

Table 1. Input cost rates per machine per quarter.

	Quarterly costs in thousands of dollars				
Input	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
Labor	\$5–\$7.5	\$5–\$7.5	\$5–\$7.5	\$5–\$7.5	\$5–\$7.5
Shears	\$7–\$10	\$14–\$19	\$14–\$19	\$7–\$10	\$9–\$12
Presses	\$6–\$9.6	\$16–\$21	\$12–\$15	\$6–\$9.6	\$5–\$9
Lathes	\$7–\$11	\$7–\$11	\$7–\$11	\$7–\$11	\$7–\$11

Table 2. AR constraints.

	$i=2$	$i=3$	$i=4$
$K=1$	$0.93 \leqslant \frac{v_2}{v_1} \leqslant 2$	$0.8 \leqslant \frac{v_3}{v_1} \leqslant 1.92$	$0.93 \leqslant \frac{v_4}{v_1} \leqslant 2.2$
$K=2$	$1.87 \leqslant \frac{v_2}{v_1} \leqslant 3.8$	$2.13 \leqslant \frac{v_3}{v_1} \leqslant 4.2$	$0.93 \leqslant \frac{v_4}{v_1} \leqslant 2.2$
$K=3$	$1.87 \leqslant \frac{v_2}{v_1} \leqslant 3.8$	$1.6 \leqslant \frac{v_3}{v_1} \leqslant 3$	$0.93 \leqslant \frac{v_4}{v_1} \leqslant 2.2$
$K=4$	$0.93 \leqslant \frac{v_2}{v_1} \leqslant 2$	$0.8 \leqslant \frac{v_3}{v_1} \leqslant 1.92$	$0.93 \leqslant \frac{v_4}{v_1} \leqslant 2.2$
$K=5$	$1.2 \leqslant \frac{v_2}{v_1} \leqslant 2.4$	$0.67 \leqslant \frac{v_3}{v_1} \leqslant 1.8$	$0.93 \leqslant \frac{v_4}{v_1} \leqslant 2.2$

Table 1 can be used to set AR constraints corresponding to the various pairs of multipliers. See Table 2 for the full set. Note that the lower bounds on all constraints $c_{iL}^k \leqslant v_i/v_n \leqslant c_{iU}^k$ can be expressed as the ratio of the lowest value v_i can take divided by the highest value v_n can assume. Similarly, the upper bounds are defined as the ratio of the highest value v_i can take divided by the lowest value taken by v_n. For example, given that the range for labor cost is \$5–\$7.5 and the range for shears is \$7–\$10 in the case of $k = 1$, the AR constraints corresponding to v_2 and v_1 are expressed as $(7/7.5) \leqslant (v_2/v_1) \leqslant (10/5)$.

All constraints are expressed in terms of labor (v_1), which has been chosen as the numeraire. Refer to Appendix B for a detailed discussion on generating a single set of AR constraints for each pair of input multipliers.

Following the methodology presented in §3, Model (11) is applied to the data of Appendix C, Tables C.1 and C.2. Recall that the purpose of solving this stage 1 problem is to facilitate an apportioning of the inputs to the subunits that make up the DMU. To bound the values of α so that a representative apportioning occurs, survey data from a sample of the plants suggested the following ranges:
N1: (0.15, 0.80)
N2: (0.10, 0.60)
N3: (0.20, 0.90)
N4: (0.20, 0.90).

It is noted that the ranges vary according to the DMU subgroup N_p and are related to the number of subunits K comprising the subgroup. Specifically, the more subunits that N_p contains, the narrower are the ranges. Recall that

$$L_{N_1} = \{R_1, R_2, R_3\}, \quad L_{N_2} = \{R_2, R_3, R_4, R_5\},$$
$$L_{N_3} = \{R_3, R_5\}, \quad L_{N_4} = \{R_1, R_3\}.$$

For example, since DMU subgroup N_2 contains four subunits, a minimum of 10% and a maximum of 60% of each input can be assigned to any subunit. In the case of N_3, however, which contains only two subunits, the alpha range is wider.

Applying Model (11), the $\hat{\alpha}_{iR_kp}$ for each DMU j_o in N_p have been derived. The results are displayed in Appendix C, Tables C.3–C.7. Recall that $\hat{\alpha}_{iR_kp}$ values are used to adjust the corresponding data in each DMU subgroup M_{R_k}, in preparation for the subunit analysis. Specifically, using the appropriately adjusted data, Model (12) is applied to each DMU in M_{R_k}, resulting in the subunit scores displayed in Appendix C, Table C.9. To derive an overall efficiency score for each DMU j_o in N_p, the relevant subunit scores are combined using the weights $W_{R_k j_o}$, as per Appendix C, Table C.8. The resulting overall scores are presented along with their relevant subunit scores in Appendix C, Table C.9.

It is noted that none of the DMUs are technically efficient. Recall that a DMU can be efficient only if all subunits for that DMU are efficient as well. However, within each subgroup R_k at least one of the (sub) DMUs in M_{R_k} is efficient.

To demonstrate the degree of sensitivity of the overall efficiency scores (as per the rightmost column in Appendix C, Table C.9) to the choice of alpha ranges, the above analysis was repeated, but with two new sets of alpha ranges. A summary of the results is as follows:

Scenario	Lower and upper limits on $\hat{\alpha}_{iR_kp}$	Average absolute change in overall efficiency scores
1	(0.10, 0.80)	0.12219
2	(0.05, 0.90)	0.33868

Based on this, it would appear that very wide ranges such as those given as scenario 2 result in substantial swings in the overall efficiency scores when compared with base results as described above. Somewhat tighter ranges such as those in Scenario 1 significantly decrease the variation in efficiency scores vis-à-vis the base results.

To complete the analysis of this section, it is worth comparing the efficiency results obtained using our model with what would have occurred had conventional DEA analysis been carried out, by simply inserting zeros in the data for any missing outputs. Two levels of analysis were conducted, namely one without any AR constraints, and one with the AR constraints applied. The results are displayed in Appendix C, Table C.10. It is noted that having replaced all blank spaces with zeros, a significant number of DMUs are rendered technically efficient. In the non-AR versions of our model and the conventional DEA model, we note that there are 3 efficient DMUs in the former versus 33 in the latter. The existence of the very large number of efficient units (33) with the conventional model is partially due to the large number of outputs and inputs involved, as compared to the total number of DMUs. A somewhat more realistic set of scores arises with the conventional model in the presence of AR constraints, where only 17 of the DMUs are efficient. Specifically, 17 of the 47 DMUs have a score of 100%, and another 8 have scores at the level of 90% or above. Arguably, part of the problem as well is that the absence of outputs in the various DMUs may be providing the opportunity to DMUs in any given subset N_p to negate the influence of other DMUs that are in different subsets.

7. Conclusions

This paper has examined efficiency measurement in a setting where decision-making units are nonhomogeneous. This environment violates the usual assumption in DEA that DMUs are all in the same "business," meaning that each DMU produces some amount of each output in a given output bundle, albeit in differing amounts from one DMU to the next. The problem of "missing" outputs has been addressed in the literature, but only in the context that either the missing value exists, but is not available to the analyst, or that the missing item is a quantity that the DMU intended to produce (and resources were expended in an effort to do so), but for whatever reason none was actually created. In this case, the value assigned to that output is legitimately taken to be *zero*.

Herein, we argue that in many situations the output mix can differ substantially from one DMU to another, meaning that the usual assumption of homogeneity does not hold, and therefore the DMUs involved are not directly comparable. Substituting zero or some other computed value when an output is missing, as a means of rendering DMUs "comparable," appears to be ad hoc, and fails to properly address the efficiency evaluation problem in a direct way. To address this apparent gap in the DEA literature, we present a DEA-like methodology that views the DMU as consisting of a set of business subunits. The overall efficiency of a DMU is then taken to be a weighted average (convex combination) of the efficiency scores for the subgroups that make up that DMU. The methodology is applied to the efficiency evaluation problem for a set of steel fabrication plants.

One criticism of this approach is that it presumes that the DMU can be viewed as being the sum of its parts, meaning that economies or diseconomies of scope are assumed to be nonexistent. In cases where this assumption is violated, our approach may fail to accurately capture the performance of the DMU. See, for example, Pulley and Braunstein (1992). Capturing economies of scope is difficult in settings where one does not have the benefit of observing an entity operating by itself, as well as in a mode where it is combined with other entities. As noted from the literature, data on mergers and acquisitions can be a way in which one might reasonably examine an entity in both states. There is also the added difficulty of separating economies of scope from economies of scale. Is the increase in output of a given product, when a new product is added to the mix, due to scope or simply a result of increased size of the operation (scale)? Current research by the authors is aimed at investigating several firms in the same industry (much like the data set herein), where product lines have been added over time. In this way, it is hoped that impact of economies of scope can be studied.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/opre.2013.1173.

Acknowledgments

The authors express their gratitude to two anonymous referees and an associate editor who made valuable suggestions for improvement in earlier versions of this manuscript. Supported under the Natural Sciences and Engineering Research Council [Grant A8966].

References

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* 2(6):429–444.

Cook WD, Hababou M (2001) Sales performance measurement in bank branches. *OMEGA* 29(4):299–307.

Cook WD, Zhu J (2011) Multiple variable proportionality in DEA. *Oper. Res.* 59(4):1024–1032.

Cook WD, Hababou M, Tuenter H (2000) Multicomponent efficiency measurement and shared inputs in data envelopment analysis: An application to sales and service performance in bank branches. *J. Production Anal.* 14(3):209–224.

Cook WD, Harrison J, Rouse P, Zhu J (2012) Relative efficiency measurement: The problem of a missing output in a subset of decision making units. *Eur. J. Oper. Res.* 220(1):79–84.

Färe R, Grosskopf S (1996) *Intertemporal Production Frontiers with Dynamic DEA* (Kluwer Academic, Boston).

Pulley LB, Braunstein YM (1992) A composite cost function for multiproduct firms with an application to economies of scope in banking. *Rev. Econom. Statist.* 74(2):221–230.

Thompson RG, Dharmapala PS, Thrall RM (1993) Importance for DEA of zeros in data, multipliers, and solutions. *J. Production Anal.* 4(4): 379–390.

Thompson RG, Langemeir LN, Lee C, Lee L, Thrall RM (1990) The role of multiplier bounds in efficiency analysis with application to Kansas farming. *J. Econometrics* 46(1–2):93–108.

Wade D. Cook is a professor of management science at the Schulich School of Business, York University, Toronto, Canada. His research interests include productivity and benchmarking, multiple-criteria decision analysis, and applications of data envelopment analysis.

Julie Harrison is a senior lecturer in management accounting at the University of Auckland Business School, Auckland, New Zealand. Her research interests include performance measurement and valuation issues, with a particular emphasis on the use of data envelopment analysis.

Raha Imanirad is a research associate at the Schulich School of Business, York University, Toronto, Canada.

Paul Rouse is an associate professor in management accounting at the University of Auckland Business School, Auckland, New Zealand. His research interests include performance measurement and data envelopment analysis, cost modelling, and revenue management.

Joe Zhu is professor of operations and industrial engineering, School of Business, Worcester Polytechnic Institute. His research interests include issues of productivity and benchmarking, and applications of data envelopment analysis.