European Journal of Operational Research 52 (1991) 1-15
North-Holland

Invited Review

Applied data envelopment analysis

A. Boussofiane, R.G. Dyson and E. Thanassoulis
Warwick Business School, University of Warwick, Coventry CV4 7AL, UK

Abstract: Data envelopment analysis (DEA) is a linear programming based technique for measuring the relative performance of organisational units where the presence of multiple inputs and outputs makes comparison difficult. This paper introduces the technique and focuses on some of the key issues that arise in applying DEA in practice.

Keywords: Efficiency, target setting, data development analysis

1. Introduction

Data envelopment analysis is an approach comparing the efficiency of organisational units such as local authority departments, schools, hospitals, shops, bank branches and similar instances where there is a relatively homogeneous set of units.

In the simplest case where a process or unit has a single input and a single output efficiency is defined simply as:

$$\text{Efficiency} = \frac{\text{output}}{\text{input}}.$$

More typically processes and organisational units have multiple incommensurate inputs and outputs and this complexity can be incorporated in an efficiency measure by defining the efficiency as:

$$\text{Efficiency} = \frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}}.$$

This definition requires a set of weights to be defined and this can be difficult, particularly if a common set of weights to be applied across the set of organisational units is sought. This problem can be resolved by arguing that individual units may have their own particular value systems and therefore may legitimately define their own peculiar set of weights. Charnes et al. (1978) propose that the efficiency of a target unit j_0 can be obtained by solving the following model:

$$\text{Max } h_0 = \frac{\sum_{r=1}^{t} u_r y_{rj_0}}{\sum_{i=1}^{m} v_i x_{ij_0}} \tag{M1}$$

subject to

$$\frac{\sum_{r=1}^{t} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, \quad j = 1, \ldots, n,$$

$$u_r, v_i \geq \varepsilon, \quad \forall r \text{ and } i,$$

where

y_{rj} = amount of output r from unit j,
x_{ij} = amount of input i to unit j,
u_r = the weight given to output r,
v_i = the weight given to input i,
n = the number of units,

Received January 1991

t = the number of outputs,
m = the number of inputs.
ε = a small positive number.

In the solution to this model the efficiency of unit j_0 is maximised subject to efficiencies of all units in the set having an upper bound of 1. The key feature of the above model is that the weights u_r and v_i are treated as unknown. They will be chosen so as to maximise the efficiency of the targeted unit j_0. The efficiency of unit j_0 will either equal 1 in which case it is efficient relative to the other units or will be less than 1 in which case the unit is inefficient. For an inefficient unit the solution identifies corresponding efficient units (i.e. efficient with the same weights) which are said to form a peer group for the inefficient unit (see also Section 3.1).

Efficiencies of all units relative to the set can be found by solving a similar model to (M1) targeting on each unit in turn. The values of the weights would of course generally differ from unit to unit and this flexibility in the choice of weights is both a weakness and a strength of this approach. It is a weakness because a judicious choice of weights may allow a unit to be efficient, but there may be concern that this has more to do with the choice of weights than any inherent efficiency. This flexibility is also a strength, however, for if a unit turns out to be inefficient even when the most favourable weights have been incorporated in its efficiency measure then this is a strong statement and in particular the argument that the weights are inappropriate is not tenable.

The DEA model (M1) is a fractional linear program but may be converted into linear form in a straightforward way so that the methods of linear programming can be applied. The linear programming version of the model is shown in model (M2):

$$\text{Max } h_0 = \sum_{r=1}^{t} u_r y_{rj0} \tag{M2}$$

subject to

$$\sum_{i=1}^{m} v_i x_{ij0} = 100,$$

$$\sum_{r=1}^{t} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \quad j = 1, \ldots, n,$$

$$-u_r \le -\varepsilon, \quad r = 1, \ldots, t,$$

$$-v_i \le -\varepsilon, \quad i = 1, \ldots, m.$$

The objective function has been linearised by recognising that in maximising a ratio it is the relative magnitudes of the numerator and the denominator that are important and not their actual values. Thus, in model (M2) the denominator has been set equal to a constant (arbitrarily set at 100) and the numerator is being maximised. (Note that any multiple of the optimal values of the weights in (M2) will be optimal in (M1).) One version of model (M2) must be solved for each unit in turn but solution of the programs is easy and computation can be made efficient by recognising that the majority of constraints are the same in each case.

Alternatively the dual of (M2) can be solved. As the primal has $t + m$ variables the dual model will have $m + t$ constraints. The primal has $n + t + m + 1$ constraints. Since n, the number of units, is usually considerably larger than $t + m$, the number of inputs and outputs, in general it will be more time consuming to solve the primal than the dual. The dual model of (M2) can be written as follows:

$$\text{Min} \quad 100 Z_0 - \varepsilon \sum_{r=1}^{t} s_r^+ - \varepsilon \sum_{i=1}^{m} s_i^- \tag{M3}$$

subject to

$$x_{ij_0} Z_0 - s_i^- - \sum_{j=1}^{n} x_{ij} \lambda_j = 0, \quad i = 1, \ldots, m,$$

$$-s_r^+ + \sum_{j=1}^{n} y_{rj} \lambda_j = y_{rj_0} \quad r = 1, \ldots, t,$$

$$\lambda_j, s_i^-, s_r^+ \ge 0 \quad \forall j, r \text{ and } i,$$

Z_0 unconstrained.

The dual need not be seen as a mere construct for computational convenience. It throws further light into the nature of the test of relative efficiency being undertaken in DEA. (M3) seeks values of λ_j to construct a composite unit, with outputs $\Sigma \lambda_j y_{rj}$, $r = 1, ., t$, and inputs $\Sigma \lambda_j x_{ij}$, $i = 1, \ldots, m$, outperforming unit j_0. The unit j_0 will be efficient if the slacks are equal to zero and Z_0 is equal to one. In other words, when it has proved impossible to construct a composite unit that outperforms j_0. Conversely, if j_0 is inefficient Z_0 will be smaller than 1 and/or slacks will be positive. The optimal values of λ_j form a composite unit outperforming j_0 and providing targets for j_0. Z_0

represents the maximum proportion of its input levels unit j_0 should be expending to secure at least its current output levels.

2. Selection of inputs and outputs

Although the DEA model avoids the problem of agreeing on a common set of weights for the inputs and outputs, it cannot of course avoid the problem of selecting which inputs and outputs should be included in the comparisons. Clearly any resource used by a unit should be included as an input. A unit will convert resources to produce outputs so that the outputs should include the amounts of products or services produced by the unit and these products or services may be produced at different levels of quality. Hence the outputs may include a range of performance and activity measures. In addition, environmental factors which may affect the production of these outputs must be identified and included in the assessment model.

Some of the practical issues in selecting and quantifying inputs and outputs can be illustrated in the context of comparing schools. Suppose that the following set of inputs and outputs is selected as shown in Table 1.

On the input side clearly a teacher in a school is a key resource and can be measured either just by the number or the number at different levels of experience or by the salary cost of the teachers. The quality of pupils at entry can also be viewed as a resource as a school is potentially adding value to the pupils. Measures of the quality of pupils at entry may be available where national testing schemes exist. If a measure of this kind is not included, then schools attracting more able pupils will automatically appear better through their higher academic achievements. The support of parents to a school can be seen as an additional resource input. Parents of higher social class, or of high levels of education, in general tend to support their children and schools more by encouragement, making sure homework is done, and perhaps direct involvement in the school through parents' associations. This factor can be classed as an environmental factor and can be included as an input where the environment is effectively providing additional resource. The final input, funds for teaching materials, is clearly a resource input to the school.

Table 1
Inputs and outputs for comparing schools

Inputs	Outputs
No. of teachers	No. of GCSE passes
Quality of pupils on entry	No. of A Level passes
Social class of parents	Standard of sport
Funds for teaching materials	Standard of music
	Employability of pupils

On the output side the academic performance of pupils is a key output and in the UK this can be measured by the number of GCSE passes (say grades A–C), which is a national examination taken at the age of 16 and by the number of A level passes which are taken at the age of 18. Clearly variations on these measures are possible by perhaps counting points for different grades and so on. Some schools may legitimately value their pupils' achievements in say sport or music and measures of these may well be appropriate outputs for the school. Again in theory the innate ability of pupils to achieve at sport or music should be allowed for at input but measures may not be readily available. Finally the employability of pupils not going on to higher education may be a further measure of performance of a school.

The schools example is thus a good illustration of the importance of including environmental factors in efficiency measurement. A school in one area may achieve better academic results than another irrespective of the quality of teaching, and a shop in one location may be favoured over another by its environment similarly. For efficiency comparisons to be useful therefore, environmental factors must be incorporated into the analysis. The environmental factor which adds resource may be included as an input whereas one that requires resource to overcome a poor environment may be included as an output. Environmental factors may be measured directly, or indirectly through the use of surrogate measures. For example the efficiency of a local authority department may be affected by social deprivation. This may be measured indirectly by the number of summonses issued to say recalcitrant community charge payers. In retailing typical environmental factors will be accessibility of the premises and competition, and again measures for these need to be incorporated.

The selection of inputs and outputs can affect the discriminating powers of DEA as the number selected needs to be small compared to the total number of units for effective discrimination. This arises due to the flexibility in the choice of weights in determining the efficiency of each individual unit. In seeking to be seen to be efficient a unit can allocate almost all its weight to a single input and output. The unit for which one particular ratio of an output to an input is highest can allocate all its weight to that ratio and appear efficient. The total number of such ratios will be the product of the number of inputs and outputs and this product is a reasonable indicator of the minimum number of efficient units. Hence with six inputs and six outputs at least 36 or so units will appear efficient, so that the total number of units in the set needs to be much greater than 36 for the method to be of any discriminatory value.

3. Managing performance using DEA

The previous sections have discussed how DEA models can be constructed to assess units for their comparative performance. This section focuses on the uses that can be made of information obtained from the DEA models. The uses discussed are some of the main ones found in applications of DEA. They are in no way exclusive or exhaustive.

One key output from the solution of the DEA model is of course the measure of the relative efficiency it yields for each unit. The relative efficiency score of a unit represents the maximum proportion of its inputs the unit should have been using, if efficient, in order to secure at least its current output levels. Alternatively, the inverse of the efficiency score is the minimum factor by which the current output levels of the unit can be multiplied for the unit to be efficient while its inputs remain at their current levels. Thus, DEA not only leads to an identification of the most and least efficient units but also to measures of the conservation of resources and/or augmentation of outputs possible.

The measures of the relative efficiencies of the units represent only the first and perhaps more obvious kind of information a DEA assessment yields. DEA can be used in a number of other ways to elaborate further on the performance of individual units and to ascertain how the units can become more efficient. Some of the further uses to which DEA can be put are the following:

- using peer groups,
- identifying efficient operating practices,
- target setting,
- identifying efficient strategies,
- monitoring efficiency changes over time, and
- resource allocation.

In the remainder of this section these uses of DEA are outlined.

3.1. Peer groups

For each inefficient unit, DEA identifies a set of corresponding efficient units said to form a peer group for the inefficient unit. Each peer unit is efficient with the inefficient unit's weights and can be identified directly from the solution to (M2). In the solution to (M3) a composite efficient unit is formed from the peer group. Peer units are associated with basic λ's.

DEA normally allows free choice of input output weights. A unit being assessed adopts a weighting structure which emphasises the particular inputs and outputs which show it in the best possible light. This can be said to identify the input output orientation of the inefficient unit and through its peer group it also identifies the subset of efficient units which could be said to have the same input output orientation as the inefficient unit as they are efficient with the same weights.

In many practical DEA assessments one needs more than the mere mathematical results to communicate to a relatively inefficient unit that their performance could improve in comparison to that of other units. Peer units can prove helpful in this respect. Merely contrasting the input/output levels of the relatively inefficient unit with those of its peer units often helps to highlight inadequacies in the performance of the relatively inefficient unit. For an illustration of this see Thanassoulis et al. (1987, Table 3). In such comparisons it often helps to scale in some way the input/output vector of peer units to make them more easily comparable to the inefficient unit. (Such scaling can be a direct output of the DEA assessment of a unit as is the case with the Warwick DEA software, 1989.)

Also given the fact that the peer units are relatively efficient and have a similar input/output orientation, they should provide examples of good operating practice for the inefficient unit to study. They may also provide suitable targets for it particularly where some peer units are of a similar size to the inefficient unit.

In using peer units in any one of the foregoing manners it is worth bearing in mind that they do not typically all have equal importance in the construction of the corresponding composite unit. It is straightforward to compute the percentage of each input/output level of the composite unit that is contributed by each peer unit. A glance at these percentages ought to reveal the extent to which any peer unit predominates in the construction of the composite unit. Such a predominant peer, if it exists, would be the chief comparator for the inefficient unit.

Finally, it must be borne in mind that the peer units identified for a given relatively inefficient unit depend on the precise DEA model used to assess it. Models (M2) and (M3) yield the same peer group when either model has a unique optimal solution. However, that peer group need not, and indeed generally will not, be the same as that obtained when solving one of the modified DEA models referred to in Section 3.3 for estimating suitable targets for an inefficient unit.

3.2. *Identifying efficient operating practices*

Identification and dissemination of good operating practices can lead to improved efficiency not only for relatively inefficient units but also for relatively efficient ones. The relatively efficient units identified are the obvious source of good operating practices. However, even amongst efficient units some are more likely to be good examples than others.

The need to discriminate between relatively efficient units in seeking good practice stems from the very essence of the DEA model which allows a unit to choose weights for the inputs and outputs so as to secure a maximum efficiency rating for itself. This can lead to a unit appearing relatively efficient by ignoring within its weighting structure all but very small subsets of its inputs and outputs. Furthermore the inputs and outputs weighed could be of secondary importance while those ignored could relate to the main function of the unit. For example in an assessment of rates departments in the United Kingdom, Thanassoulis et al. (1987) found that a certain unit appeared efficient by exclusively weighing summonses served for non payment. This was a secondary output while the number of accounts from which rates are collected which would be a primary output of a rates office was ignored. Clearly if efficient operating practices are being sought some discrimination between efficient units is necessary to ensure that relative efficiency is not simply a consequence of a totally unsupportable weighing structure.

Any one or a combination of methods can be used to discriminate between relatively efficient units. Some of the methods that can be employed are the following:
- Cross Efficiency Matrix,
- distribution of virtual inputs and outputs, and
- weight restrictions.

3.2.1. *Cross Efficiency Matrix*

A Cross Efficiency Matrix (Sexton et al., 1986) is a table which conveys information on how a unit's relative efficiency is rated by other units. Table 2 is an example of a Cross Efficiency Matrix. The entry in cell ij shows the relative efficiency of unit j with the DEA weights optimal for the target unit i. For example in Table 2 unit 1 has a relative efficiency of 1 with its own optimal weights set and relative efficiency of 0.8 with the weights optimal for unit 2.

One could compute the average of the efficiencies in each column to get a measure of how the unit associated with the column is rated by the rest of the units. A relatively efficient unit with a low such average efficiency is likely to feature in the peer groups of few inefficient units and it is likely to rely on weights dissimilar to those of the rest of the units in order to appear efficient. The converse is likely to be the case with a relatively efficient unit with a high average efficiency. Good practice is more likely to be exhibited by relatively efficient units offering high average efficiencies in their associated columns in the Cross Efficiency Matrix.

Table 2
Cross efficiency matrix – n units

Target unit	Unit 1	Unit 2	...	Unit n
1	1	0.85		0.9
2	0.8	1		0.75
3	0.92	1		1
⋮	⋮	⋮		⋮
n	1	1		1

A simple count of the frequency by which an efficient unit appears in the peer groups of inefficient units is an alternative indicator of good practice. The count indicates the extent to which a relatively efficient unit is a self evaluator or an evaluator of other units. If the count is very small in comparison to the number of units being assessed it means that the unit is largely a self evaluator. Such a unit is unlikely to offer truly efficient performance, at least not that can be readily adopted by inefficient units.

3.2.2. Distribution of virtual inputs and outputs

The virtual outputs of unit j_0 are the products of its output levels and the corresponding optimal output weights. Thus the virtual output r for unit j_0 is $u_r^* \ y_{rj_0}$ where u_r^* is the optimal value of u_r in model (M2). Virtual inputs for unit j_0 are defined in an analogous way.

Virtual input and output values convey information on the importance a unit attaches to particular inputs and outputs in order to attain its maximum efficiency rating. They can be seen as normalised weights. The levels of the actual input/output weights in the DEA model (v_i and u_r) are dependent on the scale of the corresponding input or output. In contrast, for unit j_0 the sum of its virtual outputs equals its percentage efficiency rating and so the individual virtual outputs show the contribution to that rating by each output. Virtual inputs always sum to a pre-set constant, normally 100. (See (M2).) As in the case of virtual outputs, virtual inputs show the extent to which particular inputs feature in the comparison of the one unit with others.

The inputs and outputs on which an efficient unit offers high virtual values are those which it wishes to be weighed most heavily in its comparison with other units. They thus give indications of particular areas of good practice in addition to potential overall good practice. It should be pointed out however that a DEA model may have alternative optimal solutions and corresponding virtual inputs and outputs.

Where two efficient units have contrasting patterns of virtual inputs and outputs they could be employing good operating practices in different aspects of their functions. Hence there may be scope for each unit to pick up good practice from the other and for both to improve performance.

For example in Thanassoulis et al. (1987) it is noted (Table 5) that in an assessment of rates departments 85.2% of total virtual output in the case of Brent is accounted for by summonses and distress warrants obtained. In contrast, in the case of the City of London 93.73% of total virtual output is accounted for by the value of rates collected. Both authorities were relatively efficient. If for convenience it is assumed that the virtual outputs in both cases are unique then the City of London and Brent rates departments could be said to have contrasting performances. Brent exhibits potentially good operating practices in issuing summonses and distress warrants while the City of London appears to collect efficiently money from rates payers. Potentially they could benefit from one another by exchanging information on their operating practices.

3.2.3. Weights Restrictions

An alternative way to discriminate between relatively efficient units and eliminate those relying on an inappropriate weighting structure would be to constrain the weighting structures that may be used by units. The assessment would then reveal which units are relatively efficient within the weights limits imposed. Confidence would be enhanced that such units offer genuinely efficient performance.

Some approaches to imposing weights restrictions have been suggested by Dyson and Thanassoulis (1988), Charnes et al. (1989) and Wong and Beasley (1990). They differ in the way they attempt to estimate appropriate weights restrictions. However, in all cases the ultimate result is a set of implicit or explicit weights limits to be incorporated in the DEA model solved. Once weights restrictions are imposed the DEA assessment is carried out in the usual way to yield measures of the relative efficiencies of units. All relatively efficient units must now in principle offer good operating practices. However, to the extent that there still exists a degree of flexibility on weights virtual inputs and outputs can still be used to point out the units more likely to offer good practice in a particular use of resource or area of activity.

It should be pointed out at this juncture that although weights restrictions have been introduced as a means of discriminating between otherwise relatively efficient units, that is not the only context in which such restrictions can be used. They can be used more generally to attempt to obtain efficiency measures more compatible with the perceived 'values' of inputs and outputs.

3.3. Target setting

In practical situations it is very often desired to set targets for relatively inefficient units to guide them to improved performance. Such targets provide concrete bench marks by which units can monitor their performance.

All DEA assessments yield as a by product a set of input/output levels that would render a unit relatively efficient. For example if model (M3) is used to carry out DEA assessments then where unit j_0 is relatively inefficient the following input/output levels would render it relatively efficient:

$$x'_{ij_0} = Z_0^* x_{ij_0} - s_i^{-*}, \quad i = 1, \ldots, m, \tag{1a}$$

$$y'_{rj_0} = y_{rj_0} + s_r^{+*}, \quad r = 1, \ldots, t. \tag{1b}$$

A star superscript to a variable is used to denote its optimal value. This set of targets is input orientated as the main changes are to the input levels. The DEA model solved has identified a combination of units which can secure the output levels in (1) using the target input levels in that expression. The efficient input/output levels in (1) can be readily converted to an equivalent set where input levels are not allowed to deteriorate while a maximum pro rata expansion of output levels is effected. More precisely, the output oriented efficient input/output levels would be as follows:

$$x''_{ij_0} = x_{ij0} - s_i^{-*}/Z_0^*, \quad i = 1, \ldots, m, \tag{2a}$$

$$y''_{rj_0} = \left(y_{rj0} + s_r^{+*}\right)/Z_0^*, \quad r = 1, \ldots, t. \tag{2b}$$

The relatively efficient input output levels in either (1) or (2) can be used as targets for unit j_0. In practice however these targets are only two out of an infinite set of targets that would render unit j_0 efficient. There may well exist other targets that are more desirable or indeed more meaningful for the unit. The remainder of this subsection outlines three different approaches to estimating targets for an inefficient unit other than those in (1) or (2).

Banker and Morey (1986) modified the basic DEA assessment model (model (M3)) to allow for the fact that certain inputs or alternatively certain outputs may be exogenously fixed and thus not controllable by the unit.

Where some inputs are fixed, the input oriented targets and the relative efficiency measure are based on the maximum pro rata reduction feasible to the controllable input levels. Over and above the reduction to the controllable input levels the targets also show increases in output levels necessary for efficiency. The approach also provides output orientated targets, and targets for the case where some outputs are fixed.

Thanassoulis and Dyson (1988) developed DEA models specifically designed to yield targets for a unit that are most compatible with a set of preferences over changes to individual input/output levels. Some of the models yield a measure of the relative efficiency of the unit as a by-product while other models do not. Models have been developed for estimating targets in the following cases:

- *One input or output level is given pre-emptive priority to improve.* The target input/output levels to render unit j_0 relatively efficient are determined by reducing (increasing) to its lowest (highest) level the input (output) given pre-emptive priority to improve without allowing any deterioration of the levels of the remaining inputs and outputs.
- *The unit has a general preference structure over input output changes.* The target input output levels that would render unit j_0 relatively efficient are determined in a manner that reflects the relative desirability expressed by the decision maker about changes to the levels of individual inputs and outputs.
- *The unit can specify a set of ideal target levels.* In this case targets for the unit are estimated in a two-phase process. First, a set of feasible input output levels is determined in line with the relative desirability of changes to input and output levels expressed by the user. In the second phase efficient target input and output levels are determined, if they exist, such that they dominate the input/output levels estimated in phase one.
- *Some input and/or output levels are exogenously fixed.* The models developed by Thanassoulis and Dyson (1988) can be easily modified to estimate targets for the case where some input and/or output levels are exogenously fixed. The targets estimated are in broad terms as outlined above except that preferences for changes to individual

input/output levels relate only to those levels that can be controlled by the unit.

Golany (1988) suggests an interactive approach for estimating a target set of output levels given the input levels of a unit. The aim is to enable the user to select target output levels that he or she perceives as most effective with respect to the goals of the unit concerned.

3.4. Disentangling managerial and policy efficiencies

One of the first uses of Data Envelopment Analysis was as a tool to assess the efficiencies of policies within which managers operate as distinct from the efficiencies of the managers themselves. Specifically Charnes et al. (1981) using DEA compared the Program Follow Through (PFT) and Non Follow Through (NFT) schemes for primary level school children. PFT was a large scale experiment in public education in the United States. Children on PFT came from disadvantaged backgrounds and state provision was made to remedy the effects of this at the early stages of their primary education. The experiment also involved a control group of NFT for each district that participated in PFT. Any comparison of the two programmes by merely referring to the achievements of the children of each programme is deficient in that it ascribes any inefficiencies of school management to the programmes within which those managements operate or vice versa.

DEA can be used to compare policies or programmes within which units operate in a straightforward manner. Let us assume that there are N_1, $N_2, \ldots, N_k$ units within each of $1, 2, \ldots, k$ programmes. For example in the context of assessing the branches of a bank N_1 might be the number of branches where staff of all grades are charged with promoting sales of financial products, N_2 might be branches where there is some specialisation of staff who are charged for promoting certain financial products and so on. For each set of N_i units a DEA assessment can be run in the usual manner, e.g. by solving model (M2). This assessment yields a measure of the *managerial* efficiency of each unit within programme i.

The programmes can now be compared if managerial inefficiency is notionally eliminated and all units within each programme are treated as if they operate on the efficient boundary for the programme by moving inefficient units to their target values. A new DEA assessment is carried out comprising this time the units from all programmes, i.e. $N = N_1 + N_2 + \cdots + N_k$ units in total. This assessment provides the means for comparing the programmes.

As all units are now efficient within their programme any new inefficiencies observed are attributable to the programme rather than to management. Programmes can be compared, for example by calculating the average efficiency of the units within each programme.

3.5. Efficiency changes over time

In many practical applications data is available for a unit over different periods of time. For example data for bank branches may be available on a monthly, quarterly or yearly basis. If there are n units and data exists on their input/output levels in each one of k periods then a number of analyses are possible giving alternative views of the performance of the units.

It may be meaningful to add the input/output levels over the k periods and run a single assessment of the n units. This however would not offer any views as to how the efficiency of each unit varied over time. Such variability could be the consequence of variations in staff and/or operating policies or the consequence of seasonal factors affecting different units in different ways.

An attempt to capture variations in efficiency over time could be made by treating each unit as a different one for each period of time. This would give a total of nk units to be assessed. The efficiencies obtained would give indications of the variability in efficiency of each unit over time.

Another approach to capture efficiency changes is the 'window analysis' approach. This approach is illustrated by Charnes et al. (1985). It is based on defining a 'window' of p periods. Each unit is treated as a separate one within each period and assessments are run for each window involving np units.

To illustrate the approach consider the assessment of n bank branches for which we have quarterly data covering a three year period. A window of three quarters can be defined. The first DEA assessment would then be carried out treating each branch as a different unit in each one of the first three quarters. This would give $3n$ units.

Table 3
Window analysis. Relative efficiency (%)

Quarter:	1	2	3	4	...	9	10	11	12
Branch 1									
Window 1	87.8	87.3	88.4						
Window 2		87.4	87.5	87					
			.	.	.				
Window 10							94.3	97	97.7
Branch 2									
Window 1	83.9	85.7	86.1						
		.	.	.					
⋮			.	.	.				
Branch n									
Window 1	97.9	97.3	86.4						
Window 2		89.4	88.5	86.4					
			.	.	.	.			
Window 10							100	99	98.7

The window is next 'shifted'by dropping the first quarter and bringing in data for the fourth quarter instead. A new assessment is now carried out involving as in the case of the first window $3n$ units. The process is continued until the tenth window is assessed which involves each branch as a separate unit in quarters 10, 11 and 12. The results of the window analysis can be tabulated in a manner which gives views of the efficiency of each branch either as the comparator set is altered and/or as time progresses.

For example Table 3 can be constructed as a result of the ten DEA assessments carried out in the case of the bank branches referred to above.

The three figures in each row correspond to the efficiency ratings for each branch in the window relating to the row. For example the efficiencies of branch 1 taken as a separate unit in quarters 1, 2 and 3 in the first window are 87.8%, 87.3% and 88.4% respectively.

The figures in each column give a view of the efficiency of a branch during a quarter. The efficiency values reflect the relative performance of the branch in that quarter as the comparator set of units is progressively changed. The figures across each row on the other hand indicate how the efficiency of the branch changes with time within a given window.

It is clearly up to the analyst to decide the length of the windows to be used. Windows might cover periods of time over which operating conditions are similar or where seasonal effects on performance are similar or where some other factor pertaining to the units are being assessed. Window analysis enables one to 'hold' such factor(s) constant while investigating changes in efficiency. Another important advantage of window analysis is that it effectively increases the number of units available for assessment and this increases the discriminatory power of DEA.

3.6. *DEA as a tool for resource allocation*

DEA identifies relatively efficient and inefficient units and, as noted in Section 3.3, it gives an estimate of the potential for resource conservation and/or output increases in the case of inefficient units. Both these render the method a suitable aid in allocating resources between units.

The identification of relatively efficient and inefficient units gives a first indication of the direction that the transfer of resources should in principle take. Further, knowing the potential for resource conservations or output increases at individual units gives an indication of the level of any such transfer. However, in practice resource allocation is a far more complex process than a mere implementation of the DEA results. Two important considerations need to be born in mind when contemplating resource allocations:

(a) Factors not taken into account in the DEA analysis need to be addressed in connection with the transfer of resources. For example resources at an inefficient unit may not be capable of transfer to other units or certain resources at an inefficient unit if transferred may make it even more inefficient through an inability to utilise its remaining resources efficiently. In general, when considering resource allocations the practicability of the transfer of any resources and outputs as well as higher level objectives (e.g. a policy decision to retain a unit at a minimum level irrespective of efficiency) need to be borne in mind as they do not feature in the DEA analysis.

(b) Any reallocation of resources and/or outputs leads to a change in the input/output levels of a number of units. Given the relative nature of efficiency assessment in DEA changes to input output levels of even one unit could in principle lead to changes in the relative efficiency not only of that unit but of other units as well. Hence any suggested input/output level changes make it necessary to rerun the DEA assessment to ascertain the extent to which any altered input/output levels would constitute an efficient unit.

A good illustration of the use of DEA in resource allocation and a discussion of related issues can be found in Bessent et al. (1983) where DEA was used to evaluate alternative educational programmes at a community college. There, transfers of resources were considered to both efficient as well as inefficient units. Indeed, new units were proposed with allocations of resources and estimates of their output levels. DEA was then used to estimate their efficiencies. The authors pointed out that while DEA can prove useful for resource allocation decisions it cannot answer certain questions. One important such question is one of how to distribute total organisational resources between units so that maximum output from the units collectively is obtained.

DEA can also aid resource allocation in the private sector. If the set of organisational units being assessed belongs to the profit-making sector of the economy it could be argued that the performance of the units should primarily be assessed in terms of profitability. However, the main argument against this is that the attainment of profitability can be affected by environmental factors. A profitable unit may be managed efficiently or simply enjoying favourable environment factors, whilst an unprofitable unit may be badly managed or simply be experiencing adverse conditions. If profitability is not a sufficient measure of performance in the profit-making sector, it still cannot be ignored. Both profitability and efficiency are relevant to the decision making process for the units. DEA can prove useful in determining efficiency. Units could then be assessed on an efficiency/profitability matrix analogous to the product portfolio matrix as illustrated in Figure 1.

Units located in the star quadrant are the flagship units and should provide examples of good operating practice. These units may probably be operating under favourable conditions. The sleepers are profitable, yet inefficient. Their profitability is more likely to be a consequence of a favourable environment rather than good management. They should be prime candidates for an efficiency drive leading to even greater profits. Units located in the question mark quadrant have a potential for greater efficiency and possibly greater profitability. On the other hand, units in the dog quadrant are efficiently operating but low on profitability, probably due to an unfavourable environment. In the extreme case it may be sensible to divest of these units and relocate resources to other units.

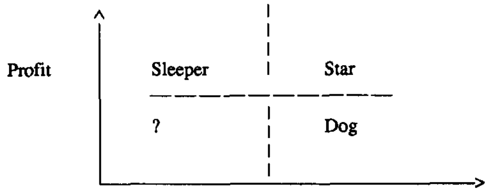

Figure 1. The efficiency/profitability matrix

As pointed out at the start of this section the foregoing uses of DEA are neither exclusive nor exhaustive. They are intended as illustrations of some important ways in which DEA can aid in the management of performance. There are a great many other uses of DEA that have been made and that can be made. Several examples can be found in Charnes et al. (1990). For example Charnes et al. (1990) present methods whereby performance of bank branches which experts agree is excellent can be used in the DEA context to discriminate between bank branches on their exposure of depositors' assets to risk. Fare et al. (1989) illustrate how DEA can be used to decompose changes in the efficiency of a unit over time between those that can be attributed to the unit and those that can be attributed to an overall change in the technology the unit is facing. Doubtless over time many new uses of DEA will be made.

4. Extensions of DEA

In Section 3 it was pointed out, among other things, how the DEA models originally built by Charnes et al. (1978) can be modified to yield information more appropriate to the needs of the assessor. Thus it was noted that restrictions could be imposed on DEA weights or DEA models can be suitably modified to yield targets compatible with desired changes to the input/output levels of the units being assessed. These were only some of the available extensions to the original DEA model. This review is concluded with a brief outline of two further extensions to the original DEA models. These extensions relax some of the assumptions underlying the original DEA models so that they may be applied in a wider set of contexts.

The two modifications outlined in this section deal with cases where the following assumptions of the original DEA models are relaxed:
- Returns to scale are constant.
- Not all input and/or output variables are continuous.

4.1. DEA models for the case where returns to scale are not constant

The original DEA model (M1) developed by Charnes et al. (1978) assumes constant returns to scale. Often this is a legitimate assumption. However, where constant returns to scale do not prevail it can be argued that units should be compared given their scale of operations or at least it would be informative to know the extent to which any inefficiency is the consequence of their scale of operations. In such cases the overall or 'aggregate' efficiency of a unit can be decomposed into its 'pure technical' and its 'scale' efficiency.

The differences between 'pure technical', 'scale' and 'aggregate' efficiencies can be illustrated by using the graph of Banker et al. (1984), reproduced as Figure 2. The figure depicts the production possibility set for the input output mix (X, Y). (X is a vector of input and Y a vector of output levels.) The line BED is the boundary of the production possibility set for the input output mix (X, Y) while x and y are scalars.

Unit A with input vector x_A X and output vector y_A Y, is inefficient. A measure of its inefficiency *for its given scale of operation, in input terms*, can be obtained if it is compared to unit B. B has the same output levels as A. The fraction x_B/x_A is '*pure technical (input)*' efficiency of A. In an analogous manner it can be seen that the factor y_C/y_A is a measure of the pure technical (output) efficiency of A. E with inputs x_E X and outputs y_E Y has the largest average productivity within the production possibility set. (The ratio y_E/x_E is maximum within the set.) E represents an 'aggregate' technically and scale efficient unit for the input output mix (X, Y). A measure of the aggregate technical and scale efficiency of A is given if it is compared with E or with 'unit N'. The latter is not within the production possibility set but in numerical terms it has the same average productivity as E. The *aggregate scale and technical efficiency* of A in comparison to N is the ratio MN/MA and it is this measure of relative ef-

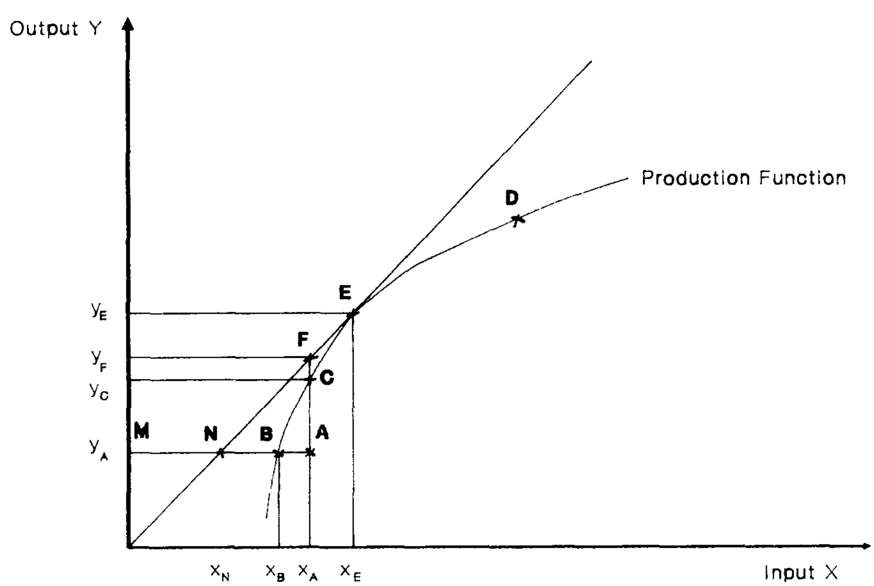

Figure 2. Technical and scale efficiencies

ficiency that the original DEA models developed by Charnes et al. (1978) yield. It can be readily shown (see Banker et al., 1984), that the aggregate efficiency of unit A equals the product of its pure technical and its scale efficiency.

Banker et al. (1984) have extended the original DEA model (M1) to assess the pure technical and scale efficiencies of units. The model that yields a measure of the pure (input) technical efficiency of unit j_0 is as follows.

$$\min h - \varepsilon\left[\sum_{i=1}^{m} s_i^+ + \sum_{r=1}^{t} s_r^-\right] \tag{M4}$$

subject to

$$hx_{ij_0} - \sum_{j=1}^{n} x_{ij}\lambda_j - s_i^- = 0, \quad i = 1, \ldots, m,$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = y_{r_0}, \quad r = 1, \ldots, t,$$

$$\sum_{j=1}^{n} \lambda_j = 1,$$

$$\lambda_j,\ s_i^+,\ s_r^- \geq 0.$$

Model (M4) differs from the original DEA model by only one constraint. That is the convexity constraint requiring that the multipliers λ_j should add up to 1. This ensures, as explained below, that (M4) yields a measure of the pure technical efficiency of unit j_0.

To see the nature of the efficiency measure yielded by (M4) note that without the convexity constraint the efficient boundary can be extrapolated from the most productive scale size for a given input/output mix. (E.g. a composite unit at point N in Figure 2 is based on an extrapolation of the average productivity at point E.) Such an extrapolated composite unit however may not be attainable for the given scale size of unit j_0. The convexity constraint ensures that the composite unit is of similar scale size as unit j_0 and it is not an extrapolation of another composite unit operating at a different scale size. Hence the efficiency measure yielded in respect of unit j_0 is its pure technical efficiency.

Model (M4) can be readily modified to assess (output) technical efficiency for a given input scale size which will reflect the extent to which outputs can be maximised while inputs do not increase. However, unlike (M3) where the efficiency measure obtained is the same whether the model used minimises inputs or maximises outputs, the technical efficiency rating obtained from models such as (M4) will in general depend upon whether input minimisation or output maximisation is chosen. Hence, in a practical context when assessments of the technical efficiency are desired, great care must be taken with the choice of an assessment model. The choice will depend upon whether inputs or outputs are controllable. In the more general case, where both inputs and outputs are controllable both measures would shed light on the performance of a unit for its scale size.

The aggregate technical and scale efficiency rating as given by (M3) is always lower than the technical efficiency rating as given by (M4). This is due to the additional (convexity) constraint in (M4). However, there will be equality between the aggregate technical and scale efficiency measure given by (M3) and the (input) technical efficiency measure given by (M4) if and only if the composite unit given by (M4), corresponding to the targeted unit, is operating at the most productive scale size. In other words, if there exists an optimal solution to (M3) such that the sum of λ_j^* is equal to 1.

Models (M3) and (M4) can be used to determine the scale efficiency of a unit. The scale efficiency of a unit (see Banker et al., 1984) is the ratio of its aggregate technical and scale efficiency to its pure technical efficiency. Model (M3) can also be used to characterize the local returns to scale for unit j_0. Unit j_0 will be operating at decreasing returns to scale if and only if the sum of λ_j^*, $j = 1, \ldots, n$, at the optimal solution to (M3) is greater than one (Banker and Morey, 1986). If the sum is lower than one unit j_0 will be operating at increasing returns to scale and if the sum is one the unit will be operating at the most productive scale size for its input output mix. (For a given input scale size a unit will be operating at decreasing returns to scale if a proportional increase of all input levels leads to a less than proportional increase in output levels. The converse case would imply the unit is operating at increasing returns to scale. In Figure 2 unit D is operating at decreasing returns to scale and unit B is operating at increasing returns to scale.)

Information as to whether a unit is operating at increasing or decreasing returns to scale can prove useful in indicating potential redistribution of resources. Resources might be transferred from units operating at decreasing returns to scale to those operating at increasing returns to scale to increase average productivity at both sets of units.

4.2. *DEA models for the case where some variables are categorical*

The original DEA model ((M2)) is based on the principle of constructing a composite unit which, if possible, would outperform the unit being targeted for assessment. The extent to which the composite unit outperforms the targeted unit gives a measure of the relative efficiency of the latter.

Any unit within the set of organisational units being assessed may be used to construct the composite unit. The latter is a linear ((M2)) or convex ((M4)) combination of a subset of the units being assessed. The subset in question consists of the peer units of the targeted unit. The construction of such composite units can lead to difficulties in certain situations.

One such situation is the case where at least one input or output variable is *categorical*. Such variables can only take a discrete set of values. E.g. in assessing university departments information on research output may only be available on an ordinal scale (good, better, excellent) or in assessing fast food outlets information on serving facilities could be categorical such as the presence or absence of drive-in facilities. In such cases a composite unit constructed as outlined above would have a flawed basis because ordinal values of categorical variables would be used as if they were measured on an *interval* scale. As a result categorical variables may be assigned meaningless values in the composite unit. E.g. if 0 stands for no drive-in facility and 1 for the presence of such a facility then a composite unit may have the value of 0.5 for the corresponding categorical variable.

In order to overcome this problem Banker and Morey (1986) have modified the original DEA model ((M2)) to ensure that the peer group of a unit may consist only of units having the same or worse values on the categorical variables than itself. This may still lead to composite units with meaningless values for categorical variables. However, it is argued that the assessment will be seen to be fairer because in essence the targeted unit is now only being compared to units operating in similar or worse conditions to those experienced by itself.

The modification of the original DEA model to cope with categorical variables is relatively simple. Consider for example the case where the categorical variable is exogenously fixed. This would be for instance the case in an assessment of schools where the social class of parents is an input. There may be three classes of parents defined. They may be 'class I', 'class II' and 'class III', where 'class III' denotes the most favourable environment in the sense that such parents best complement work by teachers.

The ideas developed by Banker and Morey (1986) would lead to the following modification of the original DEA model to cope with the categorical variable of class. Define in respect of class two zero–one variables d^1 and d^2. Then, for unit j the zero–one variables are assigned values as follows:
$d_j^1 = d_j^2 = 0$ if parents are 'class I',
$d_j^1 = 1$ and $d_j^2 = 0$ if parents are 'class II' and
$d_j^1 = d_j^2 = 1$ if parents are 'class III'.

Assuming the m-th input is the one relating to parental class the DEA model solved to assess schools is as follows:

$$\text{Min} \quad 100Z_0 - \varepsilon \sum_{r=1}^{t} s_r^+ - \varepsilon \sum_{i=1}^{m-1} s_i^- \tag{M5}$$

Subject to

$$x_{ij_0}Z_0 - s_i^- - \sum_{j=1}^{n} x_{ij}\lambda_j = 0, \quad i = 1, \ldots, m-1,$$

$$\sum_{j=1}^{n} \lambda_j d_j^1 \le d_{j_0}^1,$$

$$\sum_{j=1}^{n} \lambda_j d_j^2 \le d_{j_0}^2,$$

$$-s_r^+ + \sum_{j=1}^{n} y_{rj}\lambda_j = y_{rj_0}, \quad r = 1, \ldots, t,$$

$$d_{ji} = 0, 1, \quad \forall j, \; i = 1, 2.$$

Model (M5) differs from (M2) in the two constraints involving the zero–one variables. These constraints have replaced the constraint in (M2) relating to the m-th (categorical) input. They ensure that the peer group of school j_0 consists of schools that have parental class which is the same or worse than that enjoyed by itself.

To see this note for example that if the social class of parents for school j_0 is 'class II' then $d_{j0}^1 = 1$ and $d_{j0}^2 = 0$. The related constraints in (M5) become:

$$\sum_{j=1}^{n} \lambda_j d_j^1 \leq 1,$$

$$\sum_{j=1}^{n} \lambda_j d_j^2 \leq 0.$$

These two constraints ensure that at any feasible solution to (M5) only λ_j relating to schools that have parental classes I and II can be positive. Thus, the peer group of school j_0 will consist only of schools with parental classes I and/or II.

For more details on the treatment of categorical variables in DEA, including the treatment of controllable categorical variables see Banker and Morey (1986) and the rejoinder by Kamakura (1988).

5. Conclusion

This paper has introduced Data Envelopment Analysis (DEA) and it has highlighted some of the ways in which it can be used.

Data Envelopment Analysis is a linear programming based method for measuring the relative efficiency of organisational units. Such units, e.g. schools and bank branches, use typically a number of resources to secure a number of outputs. A key stage in a DEA assessment is the identification of the input/output variables pertaining to the units being assessed. These must reflect all resources used, outputs secured as well as the environment in which each unit operates. They must not however be excessive in number in comparison to the number of units being assessed if the method is not to lose its discriminatory power.

Apart from the measure of the relative efficiency of each unit DEA also yields other information which proves useful in gaining a better insight into the performance of each unit and in guiding units to improve their performance. A DEA assessment identifies efficient peer units for every inefficient unit. Peer units can be used to highlight the weak aspects of the performance of the corresponding inefficient unit. The input/output levels of a peer unit can also sometimes prove useful target levels for the inefficient unit. DEA yields other target input/output levels as well for each inefficient unit. The assessment model used can be manipulated to yield targets that are compatible with preferences over changes to individual input/output levels for attaining relative efficiency or to allow for the fact that certain inputs and outputs are exogenously fixed.

The efficient units DEA identifies can prove useful for providing efficient operating practices which can be disseminated to all units assessed so that they may improve their performance.

DEA yields managerial information not only in respect of individual units but also about units at the collective level. Where there are different sets of units operating under their own different policies or where a set of units has operated under different policies over time DEA can be used to ascertain the comparative efficiencies of the policies as distinct from those of the units. DEA can also be used for investigating the effects of resource transfers between units. Indeed it can be used to investigate the effects of setting up new units and/or ceasing the operations of other units.

The DEA models originally developed for comparative efficiency assessments have subsequently been modified to enable efficiency assessments to be carried out in certain special situations. The paper has outlined how DEA can be used to carry out comparative efficiency assessments in cases where returns to scale are not constant and in cases where certain inputs and/or outputs can only be measured on an ordinal scale.

There exists a very large volume of literature on DEA as witnessed by the bibliography compiled by Seiford (1990). The review in this paper is only an introductory selection and in no way exhaustive of developments in the field. Indeed there is intensive research activity in the area. For example the authors are investigating ways of building assessment models when quality of output, as well as quantity, is important. Other research areas are those of sensitivity analysis in DEA and more suitable ways of arriving at weights restrictions in DEA. Further experience with the methodology and enhancements of the models used should lead to a further expansion in the use of this management tool.

References

Banker, R.D., Charnes, A., and Cooper, W.W., (1984), "Some models for estimating technical and scale inefficiencies in data envelopment analysis", *Management Science* 30/9, 1078–1092.

Banker, R.D., and Morey, R.(1986), "Efficiency analysis for exogenously fixed inputs and outputs", *Operations Research* 34/4, 513–521.

Bessent, A., Bessent, W., Charnes, A., Cooper, W.W., and Thorogood, N., (1983), Evaluation of educational program proposals by means of data envelopment analysis", *Educational Administration Quarterly* 19/2, 82–107.

Charnes, A., Clark, T., Cooper, W.W., and Golany, B., (1985), "A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the U.S. air forces", in: R. Thompson and R.M. Thrall (eds.), *Annals of Operational Research*, 2, 95–112.

Charnes, A. and Cooper, W.W., (1990), "DEA usages and interpretations", Paper presented at IFORS 90, 12th Triennial Conference on Operations Research, Athens, Greece.

Charnes, A., Cooper, W.W. and Rhodes, E., (1981), "Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through", *Management Science* 27/6, 668–697.

Charnes, A., Cooper, W.W., and Rhodes, E., (1978), "Measuring the efficiency of decision making units", *European Journal of Operational Research* 2/6, 429–444.

Charnes A., Cooper, W.W. and Rhodes, E., (1979), "Short communication: measuring the efficiency of decision making units", *European Journal of Operational Research* 3/4, 339.

Charnes, A., Cooper, W.W., Huang, Z.M. and Sun, D.B., (1989), "DEA cone-ratio approaches for use in developing decision support systems to monitor performance in a collection of banks", Paper presented at the Conference on New Uses of DEA in Management, Austin, TX, USA.

Fare, R., Grosskopf, S., Lindgren, B. and Roos, P., (1989), "Productivity developments in Swedish hospitals: A Malmquist output index approach", Paper Presented at the Conference on New Uses of DEA in Management, Austin, Texas, USA.

Dyson, R.G. and Thanassoulis, E., (1988), "Reducing weight flexibility in data envelopment analysis", *Journal of the Operational Research Society* 39/6, 563–576.

Kamakura, W.A., (1988), "A note on the use of categorical variables in data envelopment analysis", *Management Science* 34/10, 1273–1276.

Seiford, L.M., (1990), "A bibliography of data envelopment analysis (1978–1990)", Department of Industrial Engineering and Operations Research, The University of Massachusetts, Amherst, MA, USA.

Sexton, T.R., Silkman, R.H., and Hogan, A., (1986), "Data envelopment analysis: Critique and extensions," in: Richard H. Silkman (ed.), *Measuring Efficiency: An Assessment of Data Envelopment Analysis*, Publication No. 32 in the series New Directions for Program Evaluation, A Publication of the American Evaluation Association, San Francisco, Jossey Bass, Inc.

Thanassoulis, E. and Dyson, R.G., (1988), "Setting target input output levels for relative efficiency under different priorities over individual input output improvements", Warwick Papers in Management, No. 25, Warwick Business School. To appear under the title "Estimating preferred target input output levels using data envelopment analysis", *European Journal of Operational Research.*

Thanassoulis, E., Dyson, R.G., and Foster, M.J., (1987), "Relative efficiency assessments using data envelopment analysis: An application to data on rates departments", *Journal of the Operational Research Society* 38/5, 397–411.

The Warwick DEA Software, (1989), c/o Thanassoulis, E., Warwick Business School, Warwick University, Coventry, UK.

Wong, Y.-H.B. and Beasley, J.E., (1990), "Restricting weight flexibility in data envelopment analysis", *Journal of The Operational Research Society* 41/9, 829–835.