

MULTI-ATTRIBUTE SELECTION USING DISCRETE FRONTIER PROFILING

DR. CHRIS TOFALLIS
UNIVERSITY OF HERTFORDSHIRE
BUSINESS SCHOOL
COLLEGE LANE
HATFIELD
HERTS AL10 9AB
UNITED KINGDOM

c.tofallis@herts.ac.uk

August 1997

ABSTRACT

Recent years have seen a great deal of research in the area of relative efficiency measurement in which the 'alternatives' being compared use a number of inputs to generate multiple outputs. We investigate the possibility of employing such techniques to assist with multi-attribute selection on the grounds that outputs may be viewed as criteria to be maximized and inputs as criteria to be minimized. We thus aim to find a bridge linking the fields of multi-criteria decision analysis (MCDA) and efficiency analysis.

The method which has received the most attention for dealing with multiple outputs and inputs is DEA (data envelopment analysis). We show that it is not appropriate for the selection problem. This is because it compares each alternative with an imaginary linear combination of other alternatives; whereas in a managerial context it might be feasible to adjust the inputs of an organisation to attain such a composite, in a selection problem we are not free to adjust the attribute values to do this. As a result an alternative may be classified as inefficient by DEA even though it is not dominated by any actually existing alternative.

We then look at a much less known method for efficiency measurement known as the FDH (free disposal hull) technique. This proves to be a useful link with MCDA because it is based on vector dominance. In fact with FDH the economic concept of efficiency coincides with that used in MCDA i.e. Pareto-efficiency or non-dominance. This makes FDH more suitable than DEA for the selection problem.

However, when there are many criteria, vector dominance results in many of the alternatives appearing to be non-dominated and so a lack of discrimination hinders the selection decision. This arises because all the criteria are being treated together i.e. we are looking at the problem in a space of many dimensions. In addition this will prevent us from seeing the strengths and weaknesses of each alternative because FDH (like DEA) provides a single score for each alternative.

We deal with these difficulties by grouping together related criteria and carrying out an FDH computation for each group of criteria. Since each group is a subset of all the criteria the dimensionality is reduced and so we have the benefit that discrimination is increased.

Data Envelopment Analysis

In this paper we are concerned with the problem of selecting one or more alternatives from a limited set of alternatives. For each alternative we are given scores on a number of attributes or criteria, and this data is to be used in reaching a decision. We are particularly interested in exploring the methods originally developed for efficiency measurement in the fields of management science and economics.

Data envelopment analysis (DEA) is a technique for evaluating the relative efficiency of decision making units (DMUs: e.g. firms, departments, branches etc.). In its simplest form it calculates an efficiency score as the ratio of the sum of weighted outputs to the sum of weighted inputs. It avoids the thorny problem of choosing a common set of weights by allowing each DMU to have its own personal set of weights. The weights are chosen so as to maximise the score of the DMU being evaluated subject to the condition that those weights do not cause any other DMU to exceed a score of 100%. This is achieved by solving a linear programme for each DMU. Geometrically this corresponds to setting up an efficient frontier made up of piecewise linear segments or facets which join together adjacent DMUs classed as efficient (score of 100%). The score for any inefficient unit P can be obtained as the ratio OP/OP' , where OP is the distance of the unit from the origin (O), and OP' is the distance to the frontier measured along a ray that passes through P. In general, the point P' is an imaginary alternative, it is a linear combination of other efficient units. For normal management applications of DEA this would be considered to be a feasible or achievable target point in the sense that the levels of the inputs and/or outputs could conceivably be adjusted until that point is reached. For the purpose of multi-criteria selection however, this is not appropriate because we normally do not have access to alternatives which are linear combinations of the existing alternatives. In other words DEA would be making a comparison with an alternative which does not exist.

Despite this there have been numerous applications of DEA for multi-attribute selection. The attributes where more is better are treated as outputs (\mathbf{y}) and those where more is worse are treated as inputs (\mathbf{x}). Perhaps the first application was for the location of a very large particle accelerator (Thompson et al 1986) where there were six possible sites to choose from. DEA found all six to be efficient and so additional conditions had to be imposed before a winner could be identified. More recent applications include comparing products (Doyle and Green 1991), locating a power station (Doyle and Green 1993, Tofallis 1996), technology selection (robots) (Khouja 1995), choosing a flexible manufacturing system (Shang and Sueyoshi 1995), and buying a used car (Papagiou et al 1997).

Profiling

A further difficulty with DEA is that it combines all the criteria into a single score, this makes it difficult to see the strengths and weaknesses that each alternative possesses. For instance a DEA score of 100% may be achieved merely by having the highest ratio of any one of the outputs to any one of the inputs, irrespective of how poor the scores are on all other criteria, and irrespective of whether there is any causal relationship between those two quantities. Tofallis (1996 and 1997) responded to these difficulties by introducing the simple notion of a profile for each alternative rather than a single score which attempts to summarise all of the data we have. This is achieved by partitioning the attributes into sets of related criteria and applying DEA to each such group separately. The profile then consists of a score for each group of attributes. It then becomes more difficult to achieve 100% scores because

The final version of this paper appeared in *Advances in Decision Analysis*, edited by N. Meskens and M. Roubens, (Kluwer Academic, 1999)

this would need to be achieved under every group. This is much more realistic because it is normally the case that any given alternative will have both strong points and drawbacks; profiling helps us see these more clearly than conventional DEA.

Discrete Frontiers: FDH

We have observed that DEA is not appropriate for multi-attribute selection problems because it compares alternatives with imaginary alternatives which are weighted combinations of the non-dominated set. What we require is a method which only compares alternatives with each other. This means that the 'efficient frontier' will no longer be continuous because linear combinations of efficient points are no longer permitted, hence we have instead a discrete set of non-dominated points which are no longer connected by line segments or facets. Figure 1 illustrates the difference between DEA and the proposed approach which has been called FDH (free disposal hull). In this figure y_1 and y_2 are attributes of the 'more is better' type. The DEA efficient frontier is represented by the dotted line segments which provide a convex envelope containing or enclosing all the observations/alternatives; alternatives C and E are behind this frontier and are classed as inefficient or dominated under DEA. The full lines in the diagram represent the FDH envelope which is not convex. We note that alternative C is now classed as non-dominated; it will generally be the case that FDH will result in more observations being classed as efficient than under DEA due to the envelope moving inwards towards the data points. If a short list for further scrutiny were to be produced by selecting the DEA efficient points then C would be excluded from further consideration even though it is not dominated by any other alternative. Now consider the possibility that alternative B is felt to have an unacceptably low score on attribute y_1 and D is similarly viewed with respect to y_2 ; we then have two points which each score well on one of the attributes but are deemed poor on the other. Point C might provide a happy medium. Hence FDH has the advantage of highlighting more of the well balanced alternatives as well as all those provided by DEA. Of course this will result in a longer short list.

Under FDH the efficient or non-dominated set of alternatives is established by first composing for each alternative a vector $\mathbf{v} = (\mathbf{y}, -\mathbf{x})$ which contains the attribute values. One then compares these vectors with each other, if $\mathbf{v}_1 \leq \mathbf{v}_2$ with $\mathbf{v}_1 \neq \mathbf{v}_2$ then \mathbf{v}_1 is said to be dominated by \mathbf{v}_2 (and can be excluded from further comparisons). This means that no component of \mathbf{v}_2 is worse than the corresponding component of \mathbf{v}_1 , and at least one component of \mathbf{v}_2 is better than the corresponding component of \mathbf{v}_1 . Hence a process of exclusion is used to arrive at the efficient set. This is precisely the concept of vector dominance or Pareto optimality used in multi-criteria decision analysis and so we have a link between this field and that of efficiency analysis.

It is worth noting that the FDH efficient set of alternatives can also be obtained using integer linear programming. Although computationally this would be more time consuming, it provides an interesting indication of an important break with the ideas of DEA. In DEA the evaluations are carried out using linear programming. Each linear programming problem has an equivalent dual problem which provides an alternative interpretation for what is being achieved. Suppose we formulate the original (the so-called primal) problem as the one which attempts to find a linear combination of other alternatives which is at least as good the one being evaluated in every attribute. The corresponding dual maximises the ratio of weighted outputs to weighted inputs. It is the primal, not the dual, that is adapted for FDH by imposing an integer restriction so that linear combinations are not permitted, only comparisons with individual observations. Integer linear programmes do not have equivalent duals, so we have lost contact with the ratio formulation of DEA. For multi-attribute selection problems this is

probably a good thing because otherwise we would need to justify the use of the aggregating function $(a_0 + \sum a_i y_i) / \sum b_j x_j$ which the dual is based on.

There is still a need for using the profiling approach with FDH because otherwise an alternative which had the best score on just one of the attributes would be non-dominated, irrespective of how poorly it rated on all other criteria. At least with profiling this can only occur within each group i.e. having the best score on one attribute will only make that alternative non-dominated within the category that contains that attribute, it will not be compensated in any way in the other groups of attributes.

Illustrative Application

In this section we illustrate our approach by applying it to a real data set of proposed sites for the location of a freeway interchange. There are fifteen criteria but only ten alternative locations. The data (Table 1) is taken from Lin and Teng (1992). We list below the criteria, indicating by 'max' those where these authors believed a higher value makes the site more desirable and 'min' for the reverse. It is not our intention to discuss the propriety of choosing this particular set of variables as we are merely using it for illustrative purposes.

C1(Min) Construction cost.

C2(Min) Land acquisition cost.

C3(Min) Political difficulty of land acquisition.

C4(Max) Balancing regional development.

C5(Max) Promotion of industrial development.

C6(Max) Promotion of tourism.

C7(Max) Existence of local government facilities.

C8(Max) Local roadway coordination.

C9(Max) Passenger demand.

C10(Max) Freight demand.

C11(Min) Average travel time.

C12(Min) Standard deviation of travel time.

C13(Min) Level of service of freeway near candidate location.

C14(Min) Noise impact, measured by the number of residents within 2 kilometres.

C15(Min) Air pollution impact, measured by the number of residents within 2 kilometres.

If conventional DEA is applied to all of these attributes one finds that all sites except R2 are 100% efficient. We clearly have inadequate discrimination to help us make any sort of choice. There is thus good motivation to apply the profiling approach.

We shall follow Lin and Teng in their categorization of the above into four groups:

Economy: C1 and C2

Socio-economic Development: C3-C7

Transportation: C8-C13

Environment: C14 and C15

Our profile of each candidate site will thus consist of a score for each of these categories. Under 'economy' it will suffice to simply add together the construction cost and the land acquisition cost. Under 'environment' the two impact variables are in fact being measured by the same quantity so we can use the value of this quantity without further analysis. For each of the remaining two categories we shall apply DEA and FDH and compare the results. Specifically, for the Transport category we used criteria C8, C9, C10 as y-variables, and C11, C12 and C13 as x-variables. For the Socio-economic category we used C3 as the x-variable and C4, C5, C6, and C7 as y-variables. Table 2 displays the results for

DEA (reproduced from Tofallis 1997). The values under the Economy and Environment headings are of the type 'less is better'; we have included the rankings in brackets to make the table easier to use.

SITE	ECONOMY: COST (& RANK)	SOCIOECONOMIC DEVELOPMENT (DEA SCORE %)	TRANSPORT (DEA SCORE %)	ENVIRONMENT IMPACT (&RANK)
R2	45 (9)	14	51	2.5 (8)
R3	44 (8)	21	92	3.1 (9)
R4	36 (4)	75	100	1.8 (5)
R6	46 (10)	54	85	4.2 (10)
R8	25 (2)	77	84	1.6 (4)
R9	20 (1)	96	98	1.4 (3)
R11	37 (5)	66	100	1.1 (1)
R12	42 (7)	44	100	2.4 (7)
R13	32 (3)	100	100	1.3 (2)
R14	40 (6)	67	86	2.0 (6)

Table 2: DEA profile scores.

If we restrict attention to those sites which have DEA scores of at least 80% on both of the categories assessed in this way, we find we are left with sites R9 and R13. In fact neither of these has a DEA score below 96% and site R13 scores 100% on both. Site R9 is ranked top under economy and third on environment. Site R13 is ranked second on environment and third on economy. Comparing these two locations we can say that site R13 is slightly superior in three out of four categories, but site R9 is cheaper. All that remains is for the decision makers to decide whether the superiority of site R13 is worth the extra cost.

Now we turn to the FDH results: under the transportation category all except site R2 are non-dominated. This highlights the point made earlier that the number of efficient units will generally be more numerous according to FDH as compared to DEA. For the socio-economic category we have three additional sites appearing efficient, making four in total (R4, R6, R9, R13). Recall that no analysis was carried out on the Environmental and Cost categories and so those columns in Table 2 still apply. We might therefore reject R4 for ranking fifth on environmental impact and reject R6 for coming last on both economy and environmental impact. This leaves R9 and R13 as our short list, just as before.

What our profiling approach has achieved is to reduce the eleven socio-economic and transport criteria to just two. This has simplified the problem without resorting to subjective judgement regarding the relative importance of the criteria. Whilst in this case there is no clear winner, we have at least made the problem more manageable. A short list of candidate sites can be generated.

It is worth comparing our results with those of Lin and Teng (1992) who used Saaty's analytical hierarchy process (AHP) based on weights obtained from ten experts or affected parties. They too found sites R9 and R13 to be the most preferred with overall scores which were quite close: 0.79 and 0.82 respectively, with the third choice having a noticeably lower score of 0.65. An important difference between our approach and AHP is that we delay the application of value judgements until the analysis has extracted useful information from the raw data, allowing the strongest candidates to be more easily discernible.

Conclusion and Outstanding Issues

We have shown that when a selection is involved from a limited number of alternatives the FDH technique is superior to DEA for the following reasons:

- FDH compares alternatives with each other, not with fictitious composites as with DEA
- FDH does not make any assumptions about the shape of the efficient frontier. DEA assumes it is piece-wise linear and convex.
- There is a direct link between FDH and multi-criteria decision analysis in the use of the vector dominance (Pareto optimal) concept.

The problem of having too many efficient alternatives under DEA is made worse with FDH. The profiling approach helps overcome this by treating a few attributes at a time and hence one also has the advantage of an indication of performance in each group or category instead of a single score which attempts to somehow summarise all attributes. Of course that opens up another question: How are the attributes to be grouped? In some ways this is akin to the question of which attributes are deemed to be relevant. A safe strategy for dealing with both is to investigate more than one set of attributes and more than one set of categories.

We have skirted around one issue in our presentation. As we had a number of efficient alternatives (four) which appeared in both of the categories to which FDH was applied it was not necessary to calculate FDH scores. This is because the remaining alternatives are clearly dominated. One can imagine however instances where the efficient alternatives in one category do not appear efficient in another category. In that case it would be necessary to look at ways of attaching scores using FDH, this is not without its difficulties as slacks are an even more serious problem than under DEA. See the next section for references to the literature.

Further Reading

A number of books have been published on DEA, Norman and Stoker (1991) provides a practical and readable introduction with a number of case studies, it contains a useful twenty page appendix which explains the mathematical formulations of the most widely used models. Charnes et al (1994) is a more demanding work by the originators of the subject, it also includes numerous chapters of contributed applications by other researchers, as well as a DEA bibliography consisting of 472 references covering the period 1978-1992.

The use of profiling with DEA appears in Tofallis (1996a, 1996b, 1997).

The FDH literature is much smaller and more recent than DEA. Good introductions are those of Tulkens (1993) and Fried et al (1996). In the present paper we have not discussed the calculation of scores under FDH, as with DEA there are a number of models for doing so and these can be found in De Borger et al (1995). An interesting application is Athanassopoulos and Storbeck (1992) who use FDH on a two attribute problem requiring the

The final version of this paper appeared in *Advances in Decision Analysis*, edited by N. Meskens and M. Roubens, (Kluwer Academic, 1999)

selection of two out of eight social service community centres for closure. This is somewhat unusual in that one is specifically looking for worst performance rather than best.

References

- Athanassopoulos, A.D. and Storbeck J.E. (1992). Convex versus non-convex models for spatial efficiency. Warwick Business School Research Paper no. 58.
- Charnes A, Cooper W, Lewin AY, Seiford LM (1994). *Data Envelopment Analysis: Theory, Methodology and Applications*. Kluwer Academic, Boston.
- De Borger, B., Ferrier, G.D., and Kerstens, K. (1995). The choice of a technical efficiency measure on the FDH reference technology: a comparison using U.S. banking data. Report 95/315 Universitaire Faculteiten St.-Ignatius, Antwerp.
- Doyle J and Green RH (1991). Comparing products using data envelopment analysis. *Omega* **19**(6), 631-8.
- Doyle J and Green RH (1993). Data envelopment analysis and multiple criteria decision making. *Omega* **21**, 713-715.
- Fried, H.O., Lovell, C.A.K., and Turner, J.A. (1996). An analysis of the performance of university-affiliated credit unions. *Computers and Operations Research* **23**(4), 375-384.
- Khouja, M. (1995). The use of data envelopment analysis for technology selection. *Computers and Industrial Engineering* **28**(1), 123-132.
- Lin C-Y and Teng J-Y (1992). Multicriteria decision making for selecting freeway interchange locations in Taiwan. In *Multiple Criteria Decision Making (Proceedings of the 9th International Conference)* ed. A. Goicoechea, L. Duckstein, S. Zionts.
- Norman, M, and Stoker, B. (1991). *Data Envelopment Analysis: The Assessment of Performance*. John Wiley and Sons, Chichester and N.Y.
- Papagapiou, A., Mingers, J. and Thanassoulis E. (1997). Would you buy a used car with DEA? *OR Insight* **10**(1), 13-19.
- Shang, J. and Sueyoshi, T, (1995). A unified framework for the selection of a flexible manufacturing system. *EJOR* **85**(2), 297-315.
- Thompson RG , Singleton FD, Thrall RM, and Smith BA (1986). Comparative site evaluation for locating a high energy physics lab in Texas. *Interfaces* **16**(6), 35-49.
- Tofallis, C. (1996a) Resource efficiency profiling. In *Multi-Objective and Goal Programming: Theories and Applications* ed. Tamiz M. Springer-Verlag.
- Tofallis, C. (1996b) Improving discernment in DEA using profiling. *Omega* **24**(3), 361-364.
- Tofallis, C. (1997) Input efficiency profiling: an application to airlines. *Computers and Operations Research* **24**(3), 253-258.
- The final version of this paper appeared in *Advances in Decision Analysis, edited by N. Meskens and M. Roubens*, (Kluwer Academic, 1999)

Tofallis C (1997) Multi-criteria site selection using DEA profiling. *Studies in Locational Analysis* (forthcoming 1997).

Tulkens H. (1992). On FDH efficiency analysis: some methodological issues and applications to retail banking, courts and urban transit. *J. of Productivity Analysis* **4**, 183-210.