DATA ENVELOPMENT ANALYSIS: AN EXPOSITION AND CRITIQUE

Harvey Goldstein

Institute of Education, University of London, 20 Bedford Way, London WC1H 0AL

Abstract One of the difficulties with the use of Data Envelopment Analysis (DEA) has been the relative obscurity of the mathematical techniques with which it operates. This paper outlines the nature of these techniques and by using a simple example demonstrates their inadequacy for the measurement of school performance. It also argues against the use of aggregate level data in studies of school effectiveness and suggests that the only secure basis for such research is provided by multilevel models which use measurements made on individual students.

Introduction

Applications of Data Envelopment Analysis (DEA) to education arose from attempts to define measures of 'efficiency' which could be applied to non-commercial enterprises such as schools and hospitals (Charnes, Cooper & Rhodes, 1978). In essence, efficiency is defined as a weighted sum of 'outputs' divided by a (differently) weighted sum of inputs, with the weights estimated from the data. Each unit, such as a school or a hospital, is assumed to have a measurement on each output and each input. In the case of schools the outputs might consist, for example, of average examination result or attendance rate over a period of time, with inputs such as expenditure per pupil or average intake test score.

DEA was originally developed for other contexts. In industrial processes, for example, where there are several kinds of input and output, a definition of efficiency may be required and typically this will be based on costs. In this situation, where each input and output can be measured in monetary terms, a reasonable definition of efficiency can be based upon a ratio of cost functions. Techniques such as DEA then attempt to find those combinations of inputs and outputs which will maximise this ratio.

A distinction is immediately apparent between DEA and the class of models based on linear regression. These latter relate an output measure to a set of inputs and then interpret the differences between the *predicted* outputs based on the estimated relationship, and the *actual* observed output measurements, as measures of efficiency. Such differences are the usual 'residuals' in a regression analysis. In DEA there is no attempt to take account of the nature of any relationships between output and input, rather it proceeds to *define* efficiency as a ratio. It

0950–0790/90/01 0017–04$02.50 0
EVALUATION AND RESEARCH IN EDUCATION

Vol. 4, No. 1, 1990

does, however, use data aggregated to the level of the unit as do many of the analyses based on regression techniques (Gray & Jesson, 1987; Woodhouse & Goldstein 1988).

In the next section I outline the DEA procedure and discuss its interpretations using simple examples. The following section presents a critique of the procedure, and the final section discusses alternatives.

The DEA Model

A formal statement of this model is as follows. We denote the rth output of unit j by y_{rj} and the ith input of unit j by x_{ij} for the jth unit we shall require weights u_{rj}, v_{ij} as follows. There are p units. For the jth unit, define the kth ratio

$$h_{jk} = \frac{\sum_r u_{rj}\, y_{rk}}{\sum_i v_{ij}\, x_{ik}} \qquad k = 1,\ldots\ldots,p \tag{1}$$

The required weights are those which make this ratio, for $k=j$, as large as possible relative to all other units. Before we can do this, however, we need to 'constrain' the solution. As it stands we can multiply the weights in the numerator all by any number, say a, and the weights in the denominator by any number, say b, without altering the nature of the solution. There are many ways to 'fix' the weights, for example by requiring each set to add to a constant, say 1.0. The approach of DEA is to require all the ratios to be no greater than 1.0, with a further constraint on the scale of either the numerator or denominator weights; one choice again being that either should sum to 1.0. The solution will generally be dependent on the constraints chosen, and in particular the rankings of the units will vary with the constraints chosen. This appears to be a problem which has not been studied in this context. We shall not pursue this issue, however, since there are more serious concerns.

The above procedure is carried out for each unit in turn, thus giving a set of coefficient values and a value of the ratio for each unit. These ratios are then interpreted as the relative 'efficiencies' of each unit.

To simplify matters consider the common case where there is only one output, say a school's average examination result, with several inputs. The procedure now seeks to maximise

$$h_{jk} = \frac{y_k}{\sum_i v_{ij} x_{ik}} \tag{2}$$

The weights and the input and output variables are all assumed to be positive so that consistent interpretations are possible.

We can further simplify (2) by considering the special case of just one input variable. We no longer now have a complex maximising problem, since we simply compare the ratios

$$h_k = \frac{y_k}{x_k} \tag{3}$$

for each unit. We can gain some insight into the procedure by studying this case in more detail.

Comparing Simple Ratios

In model (3) let us suppose that our output is an average examination result and our input is an average intake score for a school. Then the ratio of these two variables is defined by the DEA model as efficiency and a simple ranking of schools can be made. The interpretation, however, is not straightforward.

First, while it might be argued that the ratio of test or exam scores is a reasonable measure of efficiency, it would be more difficult to argue that a simple ratio of an exam score to the proportion of children from middle class homes is a good measure of efficiency. In typical DEA analyses with multiple input variables, these are typically a mixture of test scores and other characteristics of the school and the students attending it, so that some care needs to be paid to any interpretations.

Secondly, unlike regression models, DEA claims to pay no attention to a description of the actual relationship between output and input variables, simply concerning itself with the properties of the ratio. Nevertheless, in reality a particular underlying relationship will exist, and the nature of this relationship will determine the properties of the ratio. In general such a relationship will be complicated but the essential nature of the problem can be illustrated simply.

Suppose that the true relationship had the following very simple form

$$y_k = a + bx_k \tag{4}$$

Then from equation (3) we obtain immediately

$$h_k = \frac{a}{x_k} + b \tag{5}$$

and we see that the ratios are inversely proportional to the input scores, or intake test scores. Alternatively, suppose y, x are measured from their means μ_y, μ_x and the true relationship is

$$y_k - \mu_y = b(x_k - \mu_x) + c(x_k - \mu_x)^2 \tag{6}$$

Then

$$h_k = b + c(x_k - \mu_x) \tag{7}$$

and now the ratios are a linear function of the intake test scores. In neither case would it be appropriate to interpret the h_k as 'efficiencies'. Without knowing what the underlying model might be we can make no sensible interpretation. In the case of multiple inputs DEA effectively maximises a ratio of two linear functions one of the input variables and one of the output variables. Needless to say, in this more general case the same objections apply, although the complexity of the estimation procedures necessary may sometimes obscure this point.

Thus, even in the simplest cases we see that the DEA model used on its own will not have an unambiguous interpretation. Great care is needed in defining input (and output) variables so that proper interpretation of the term 'efficiency' can be made and, more importantly, it is clear that some information about the *relationship* between output and input variables is required. This latter requirement leads to the use of standard statistical models and we shall return to this in the next section.

Aggregate Models

The other problem with DEA, and much of the analysis based upon regression models (Gray & Jesson, 1986), is that it is based upon the use of LEA or school level data. That is, for example, whole school characteristics or the average characteristics of all the students in a school or school year.

Thus Gray & Jesson (1986) present analyses of examination results aggregated to the LEA level and seek to draw conclusions about effectiveness using the residuals from a multiple regression analysis. Woodhouse & Goldstein (1988) present a detailed critique of such models and point out their inherent instability and the consequent dangers of using them to compare aggregate units. In a similar manner, DEA analyses are open to the same objections through their use of aggregate level data. In particular their conclusions will be sensitive to assumptions made about the scale on which variables were measured, and the choice of variables used. As others have also pointed out (e.g. Aitkin & Longford, 1986) the only secure basis for comparisons of units such as schools and LEAs lies in multilevel analyses where individual pupil data are used together with information about schools and other higher level units. There is increasing evidence (Nuttall *et al.*, 1989) that school effectiveness is multidimensional. Schools are differentially effective for different groups of pupils, and this is a matter of some importance to those who wish to base policy decisions on such results. The use of a school average cannot hope to capture such subtleties.

Conclusions

While there is a certain attraction in using a simple summary of aggregate level inputs and outputs, whether in terms of a ratio as in DEA or a regression residual, these procedures are in fact surrounded by very serious problems of interpretation. There would appear to be no proper substitute for the collection and analysis of multilevel data. If such data are not available then it would be prudent to treat aggregate level data with considerable care. In particular it is difficult to see any justification for the use of DEA in studies of educational efficiency.

Acknowledgement

I am most grateful to Geoff Woodhouse for his helpful comments.

References

Aitkin, M. and Longford, N. (1986) Statistical modelling issues in school effectiveness studies. *J. Royal Statis. Soc.*, A, 149, 1–43.

Charnes, A., Cooper, W. W. and Rhodes, E. (1978) Measuring the efficiency of decision making units. *European J. of Operational Research*, 429–44.

Gray, J. and Jesson, D. (1987) Exam results and local authority league tables. In *Education and Training U.K. 1987* (pp. 33–41).

Nuttall, D. L., Goldstein, H., Prosser, R. and Rasbash, J. (1989) Differential school effectiveness. *International Journal of Educational Research*, 13, 769–76.

Woodhouse, G. and Goldstein, H. (1988) Educational performance indicators and LEA league tables. *Oxford Review of Education*, 14, 301–20.