

1-1-1997

Data Envelopment Analysis with Integral Targets

Gunter P. Sharp
Georgia Institute of Technology

Timothy R. Anderson
Portland State University, tim.anderson@pdx.edu

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/etm_fac

 Part of the [Engineering Commons](#)

Citation Details

Anderson, T.R.; Sharp, G.P., "Data envelopment analysis with integral targets [management decision making]," *Innovation in Technology Management - The Key to Global Leadership. PICMET '97: Portland International Conference on Management and Technology*, pp.794.

This Article is brought to you for free and open access. It has been accepted for inclusion in Engineering and Technology Management Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

Data Envelopment Analysis with Integral Targets

Timothy R. Anderson
Engineering Management Program
Portland State University
Portland, OR 97207-0751 USA

Gunter P. Sharp
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332 USA

Abstract-Data Envelopment Analysis (DEA) inherently assumes that the feasible targets for inputs and outputs are continuous. This paper develops and illustrates extensions to DEA that allow for integer valued inputs and outputs. It is found that DEA's implicit assumption of continuous targets can be significant in certain applications.

I. INTRODUCTION

Data Envelopment Analysis (DEA), is a technique for analyzing the relative efficiency of different decision making units (DMUs), operating in complex environments with multiple inputs and/or multiple outputs. It is a linear programming based approach which therefore has both primal and dual formulations. A brief introduction to DEA is provided in [3] which defines the basic models, variables, and extensions. A definitive reference for DEA can be found in [10].

Since the basic formulation of DEA was introduced in [12] there have been many extensions to the basic DEA model to account for different types of inputs and outputs. Two important areas of extensions in the primal formulation include restricting the ranges of multipliers, varying discretion levels over certain inputs or outputs, and categorical inputs or outputs.

The issue of restricting ranges of multipliers in the dual formulation has been well examined by for example [2, 7, 11, 14, 19, 20, 22]. Restricting ranges of multipliers in the primal formulation involves more subtle complications but has also been studied [4, 8, 23, 24]

For example, it was realized that DMUs do not always have full control over the level of each input and output. One such case could be location in a model of convenience store management performance. A store manager does not have control over where the store is located and must make the best of their current location. Therefore a model that assumed that managers should have performance targets based upon different store locations could be unrealistic. This issue led to extensions allowing nondiscretionary and partially discretionary inputs and outputs [5, 9, 15, 18].

The issue of categorical inputs and outputs has been examined by a number of researchers [6, 13, 16, 17, 21]. Categorical variables are those having a type such as "good, better, or best" rather than a quantifiable value.

In some applications, it is important to restrict the inputs or the outputs to be integers. This situation was first encountered by the authors in discussions concerning an application of gas station and convenience store operations where the number of

gasoline pumps was considered as a possible input. In this application, it was meaningless to develop a performance target for a gas station based upon a fractional number of pumps such as 6.5. On the other hand, labor may not necessarily be required to take on integer values. Full-time-equivalents could be used to include part-time employees. In situations where it might not be possible to hire people such as managers on a part-time basis, it may still be possible to have a manager split time between several stores, effectively approximating continuous values. While it may be useful to determine targets based on these continuous valued factors as a way of encouraging people to seek novel solutions to their problems, it should also be recognized that this may be an unfair comparison. In an application where an input or an output cannot be accurately modeled as having continuous values, a model that explicitly recognizes the integral nature of the application should be used.

This issue of integer valued inputs and outputs (or conversely, the issue of avoiding fractional targets) is relevant to some high technology applications. For example, certain expensive pieces of semiconductor manufacturing equipment that may be needed might not be possible to "share" so as to approximate continuous values. In this paper, we will provide new formulations that account for integer valued inputs and outputs. A numerical example will illustrate the use of this model.

II. MODEL

The key to incorporating integral inputs and outputs is to create an explicit set of virtual inputs and outputs and then to constrain these to integral values. Traditionally, the primal formulation of DEA assumes a two phase approach. In the first phase, a radial efficiency score is calculated while using no more inputs than DMU_0 and at as many outputs as DMU_0 . In the second phase, the sum of the slack inputs and outputs are maximized given that the efficiency score must be equal to the efficiency score obtained in the first phase. This is a basic preemptive goal programming approach. Rather than writing out two separate linear programs this is commonly represented by using a non-Archimedean approach. The primary goal preempts the secondary goal by the multiplication of the sum of slacks by an infinitesimal. Note that it has been demonstrated by [1] that the infinitesimal is a notational convenience and that this should not be solved by using a "small" number like 10^{-6} . With this warning, we will follow the concise notation in the manner of the non-Archimedean formulation as shown in (1).

$$\begin{aligned}
 & \min_{\theta, \lambda, s^+, s^-} \theta - \varepsilon s^+ - \varepsilon s^-, \\
 \text{s.t.} \quad & Y\lambda - Y_0 - s^+ = 0, \\
 & \theta X_0 - X\lambda - s^- = 0, \\
 & \lambda, s^+, s^- \geq 0.
 \end{aligned} \tag{1}$$

A similar approach is used to incorporate integer targets. Rather than using two steps (or phases), this approach will use three steps. The first change from the basic radial model [3] is to declare explicit sets of variables representing the target input and output levels with the additional requirement that the targets must be integers. This is illustrated by (2).

$$\begin{aligned}
 & \min_{\theta, \lambda, \hat{X}, \hat{Y}} \theta, \\
 \text{s.t.} \quad & Y\lambda \geq \hat{Y}, \\
 & \hat{Y} \geq Y_0, \\
 & X\lambda \leq \hat{X}, \\
 & \hat{X} \leq \theta X_0, \\
 & \hat{X} \text{ and } \hat{Y} \text{ are vectors of integers} \\
 & \lambda \geq 0.
 \end{aligned} \tag{2}$$

The first formulation will yield an optimal radial efficiency score but does not necessarily indicate the best integer valued target level of performance. The second phase maximizes the sum of the slacks between DMU₀ and the target while still achieving the same efficiency score. An infinitesimal is used to denote this secondary goal in formulation (3).

$$\begin{aligned}
 & \min_{\theta, \lambda, s^+, s^-, \hat{X}, \hat{Y}} \theta - \varepsilon s^+ - \varepsilon s^-, \\
 \text{s.t.} \quad & Y\lambda \geq \hat{Y}, \\
 & \hat{Y} - Y_0 - s^+ = 0, \\
 & X\lambda \leq \hat{X}, \\
 & \theta X_0 - \hat{X} - s^- = 0, \\
 & \hat{X} \text{ and } \hat{Y} \text{ are vectors of integers} \\
 & \lambda, s^+, s^- \geq 0.
 \end{aligned} \tag{3}$$

The second phase still may use “weakly efficient” DMUs in forming a target. The third phase maximizes the sum of the slacks between the targets and the virtual DMU built with λ while still achieving the earlier radial efficiency score and

slacks. Another infinitesimal, ε^2 , is used to denote this secondary goal in formulation (4).

$$\begin{aligned}
 & \min_{\theta, \lambda, s^+, s^-, \hat{X}, \hat{Y}} \theta - \varepsilon^1 s^+ - \varepsilon^1 s^- - \varepsilon^2 t^+ - \varepsilon^2 t^-, \\
 \text{s.t.} \quad & Y\lambda - \hat{Y} - t^+ = 0, \\
 & \hat{Y} - Y_0 - s^+ = 0, \\
 & X\lambda - \hat{X} + t^- = 0, \\
 & \theta X_0 - \hat{X} - s^- = 0, \\
 & \hat{X} \text{ and } \hat{Y} \text{ are vectors of integers} \\
 & \lambda, s^+, s^- \geq 0.
 \end{aligned} \tag{4}$$

III. OTHER MODELS

Similar modifications can be easily made to other primal DEA models to allow for integral inputs or outputs.

Many applications that have integral inputs and outputs will also have continuous input or output factors. Since the models developed earlier can be considered to be generalized version of the standard DEA model, accounting for mixed models can be readily accomplished by simply specifying that the specific target variable corresponding to the continuous input or output is continuous rather than being a general integer. For example, a two output model with the first output being continuous and the second one being integral would then have \hat{Y}_1 be continuous and \hat{Y}_2 be a general integer.

IV. RESULTS

A. Example with Integral Inputs and a Continuous Output

An example would help to illustrate the effects of the new integral DEA model. The first example includes two integral inputs and a continuous output. The data set used for this is given in the following Table I.

TABLE I
DATA SET WITH INTEGRAL INPUTS.

	Inputs		Output
	X ₁	X ₂	Y
P ₁	1	9	1
P ₂	3	4	1
P ₃	6	1	1
P ₄	4	9	1
P ₅	3	6	1
P ₆	7	4	1
P ₇	7	3	1
P ₈	2	9	1

First, we will consider the graphical solution with continuous inputs as given in Fig. 1. The DEA model used is a

standard input-oriented DEA model with constant returns to scale.

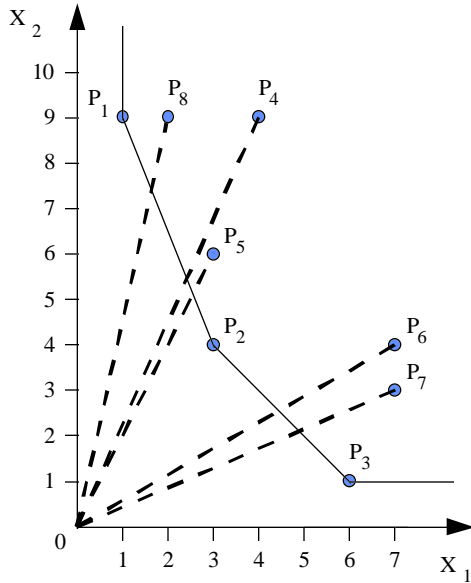


Fig. 1. Graphical analysis with continuous inputs.

Three of the DMUs lie on the efficiency frontier and the other five are radially inefficient. The numerical results are summarized in the following table.

TABLE II
NUMERICAL RESULTS WITH CONTINUOUS INPUTS.

	Inputs		Targets		Efficiency θ
	X_1	X_2	X_1	X_2	
P_1	1	9	1	9	1
P_2	3	4	3	4	1
P_3	6	1	6	1	1
P_4	4	9	2.42	5.45	0.61
P_5	3	6	2.56	5.11	0.85
P_6	7	4	4.45	2.55	0.64
P_7	7	3	4.90	2.10	0.70
P_8	2	9	1.64	7.39	0.82

Notice that the target inputs for all five of the inefficient DMUs are non-integral. Now we impose the assumption that target inputs must be integral. Fig. 2 illustrates some of these feasible targets with integral inputs. A new efficiency frontier can then be drawn by connecting the lower leftmost feasible solutions. The new efficiency frontier can be described as taking on a staircase shape.

Once the new efficiency frontier has been determined, the radial efficiency scores can be examined graphically in the same manner as in standard DEA. The radial efficiency is now calculated the same as in Section 2 with just a change in the efficiency frontier. The radial efficiency scores are easily calculated in Fig. 1. using ratios or similar triangles. The target for each DMU is more difficult to determine.

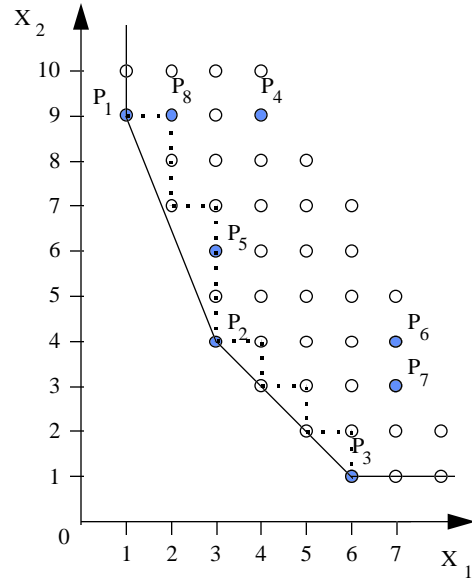


Fig. 2. Feasible targets for DMUs.

P_4 is clearly inefficient with $\theta = 3/4$. It has several possible targets including $X_1 = 3$ and $X_2 = 4, 5,$ or 6 . In general though, we will assume that slacks will be maximized, which leaves just one possible target, $X_1 = 3$ and $X_2 = 4$, or P_2 . Similarly, although P_5 is radially efficient it has extra consumption of X_2 and therefore it also has a target of P_2 . Also, P_5 is considered to “weakly efficient” but not “strongly efficient.”

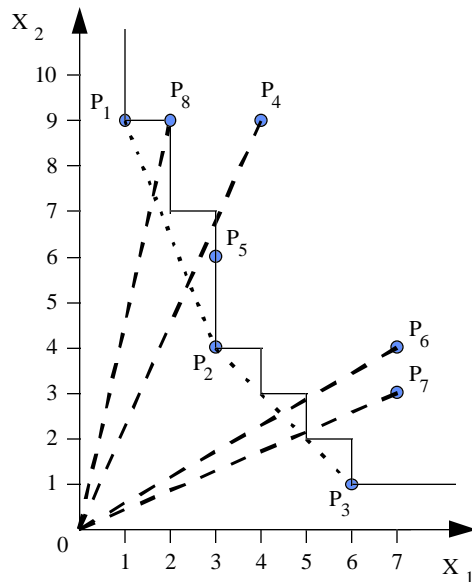


Fig. 3. Graphical analysis with integral inputs.

P_6 and P_7 are interesting for several reasons. First, they both have identical efficiency scores now of $\theta = 5/7$, but in the continuous input case they had efficiency scores different from

each other. Second, they have the same target value of $X_1 = 5$ and $X_2 = 2$. Third, the target is composed of 1/3 of P_2 and 2/3 of P_3 .

The evaluation of P_8 is more complicated. Since it lies upon the efficiency frontier, it is radially efficient and $\theta = 1$. At first glance, this might seem to indicate that P_1 should be as the best target. However, a combination of P_1 and P_2 such that $X_1 = 2$ and $X_2 = 7$ provides a higher sum of slacks and is the preferred target for P_8 .

TABLE III
NUMERICAL RESULTS WITH INTEGRAL INPUTS.

	Inputs		Targets		Efficiency θ
	X_1	X_2	X_1	X_2	
P_1	1	9	1	9	1
P_2	3	4	3	4	1
P_3	6	1	6	1	1
P_4	4	9	3	4	0.750
P_5	3	6	3	4	1
P_6	7	4	5	2	0.714
P_7	7	3	5	2	0.714
P_8	2	9	2	7	1

In this example the integrality condition had a major effect. Among the five DMUs originally inefficient, the mean efficiency increased from 0.724 to 0.836; thus the mean increase in efficiency due to the integrality constraints was 15%. The maximum increase was 23%; this occurred as P_4 rose from 0.61 to 0.75. Perhaps more importantly, two DMUs, P_5 and P_8 , joined the ranks of Debreu-Farrell efficient DMUs. Table III summarizes these results.

V. CONCLUSION

The integrality conditions significantly changed the efficiency scores. The maximum change was 23% and the mean change among inefficient DMUs was 15%. The magnitude of this effect implies that an analyst should examine the application to see if any of the inputs or outputs should be modeled as integral factors.

It is expected that integrality conditions will have a larger effect when the numbers involved are smaller and there are smaller ranges of variation in the inputs and the outputs. Further work needs to be done in examining when the integrality conditions need to be imposed.

REFERENCES

- [1] A. I. Ali, "Data envelopment analysis: computational issues," *Computers, Environment and Urban Systems*, vol. 14, pp. 157-65, 1990.
- [2] A. I. Ali, W. D. Cook, and L. M. Seiford, "Strict vs. weak ordinal relations for multipliers in data envelopment analysis," *Management Science*, vol. 37, pp. 733-8, 1991.
- [3] T. R. Anderson and K. B. Hollingsworth, "An introduction to data envelopment analysis in technology management," in *Proceedings of PICMET '97*, D. F. Kocaoglu, T. R. Anderson, K. Niwa, D. Milosevic, and M. J. Gregory, Eds. Portland: PICMET, 1997.
- [4] R. D. Banker, A. Charnes, and W. W. Cooper, "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science*, vol. 30, pp. 1078-1092, 1984.
- [5] R. D. Banker and R. C. Morey, "Efficiency analysis for exogeneously fixed inputs and outputs," *Operations Research*, vol. 34, pp. 513-21, 1986.
- [6] R. D. Banker and R. C. Morey, "The use of categorical variables in data envelopment analysis," *Management Science*, vol. 32, pp. 1613-27, 1986.
- [7] R. D. Banker and R. C. Morey, "Incorporating value judgements in efficiency analysis," in *Research in Governmental and Nonprofit Accounting*, vol. 5: JAI Press, 1989, pp. 125-163.
- [8] K.-P. Chang and Y.-Y. Guh, "Linear production functions and the data envelopment analysis," *European Journal of Operational Research*, vol. 52, pp. 215-23, 1991.
- [9] A. Charnes, W. W. Cooper, B. Golany, and L. Seiford, "Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions," *Journal of Econometrics*, vol. 30, pp. 91-107, 1985.
- [10] A. Charnes, W. W. Cooper, A. Y. Lewin, and L. M. Seiford, "Data Envelopment Analysis: Theory, Methodology and Applications," Boston: Kluwer, 1994.
- [11] A. Charnes, W. W. Cooper, Q. L. Qei, and Z. M. Huang, "Cone ratio data envelopment analysis and multi-objective programming," *International Journal of Systems Science*, vol. 20, pp. 1099-118, 1989.
- [12] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operations Research*, vol. 2, pp. 429-44, 1978.
- [13] A. Charnes, J. J. Rousseau, and J. H. Semple, "Nonarchimedean infinitesimals, transcendentals and categorical inputs in linear programming and data envelopment analysis," *International Journal of Systems Science*, vol. 23, pp. 2401-6, 1992.
- [14] R. G. Dyson and E. Thanassoulis, "Reducing weight flexibility in data envelopment analysis," *Journal of the Operational Research Society*, vol. 39, pp. 563-76, 1988.
- [15] B. Golany and Y. Roll, "Some extensions of techniques to handle non-discretionary factors in data envelopment analysis," *Journal of Productivity Analysis*, vol. 4, pp. 419-432, 1993.
- [16] B. Golany, Y. Roll, and D. Ryback, "Measuring efficiency of power plants in Israel by data envelopment analysis," *IEEE Transactions on Engineering Management*, vol. 41, pp. 291-301, 1994.
- [17] W. A. Kamakura, "A note on 'the use of categorical variables in data envelopment analysis'," *Management Science*, vol. 34, pp. 1273-6, 1988.
- [18] S. C. Ray, "Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation," *Socio-Economic Planning Sciences*, vol. 22, pp. 167-176, 1988.
- [19] Y. Roll, W. Cook, and B. Golany, "Controlling factor weights in data envelopment analysis," *IIE Transactions*, vol. 23, pp. 2-9, 1991.
- [20] Y. Roll and B. Golany, "Alternate methods of treating factor weights in DEA," *Omega*, vol. 21, pp. 99-109, 1993.
- [21] J. J. Rousseau and J. H. Semple, "Notes: Categorical outputs in data envelopment analysis," *Management Science*, vol. 39, pp. 384-386, 1993.
- [22] R. G. Thompson, F. D. Singleton, Jr., R. M. Thrall, and B. A. Smith, "Comparative site evaluations for locating a high-energy physics lab in Texas," *Interfaces*, vol. 16, pp. 35-49, 1986.
- [23] H. Tulkens, "On FDH efficiency analysis: some methodological issues and applications to retail banking, courts, and urban transit," *Journal of Productivity Analysis*, vol. 4, pp. 183-210, 1993.
- [24] D. A. Vlatsa, "Data Envelopment Analysis with Intensity Restrictions," in *School of Industrial and Systems Engineering*. Atlanta: Georgia Institute of Technology, 1995, pp. 215.