No. 827

Proposing non-uniform evaluation in social systems analysis

by

Akihiro Hashimoto

July 1999

Proposing non-uniform evaluation in social systems analysis*

Akihiro Hashimoto†

Institute of Policy and Planning Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan

Abstract

This paper considers multi-dimensionally comprehensive evaluation in social systems analysis. Although we have been employing weighted sum of evaluation items as a comprehensive evaluation tool, it cannot but be somewhat arbitrary because of the difficulty to define *a priori* weighting. We propose a new evaluation without specifying such weighting using DEA (Data Envelopment Analysis). This is a non-uniform evaluation with flexibly defined weighting that can vary by object being evaluated, making a good use of each object's feature characteristics. There seem to be many cases for which such non-uniform evaluation is rather suitable. Through some application cases, this paper demonstrates the appropriateness of the non-uniform evaluation.

Key Words and Phrases: non-uniform evaluation, social systems analysis, comprehensive evaluation, weighted sum, arbitrariness, data envelopment analysis, articles for sale, flexibly defined weights

1 Introduction

This paper addresses *evaluation* or *decision-making* in *social systems analysis*, i.e., model analysis of social systems. To evaluate or to assess is to judge the quality, performance, etc. of objects, which is used as a standard of the subject's behavior. When we put emphasis on the face of determining behavior, we may say decision-making instead of evaluation. We evaluate and then make a decision, so that there should be no decision-making without evaluation.

Evaluating objects requires considering from what point of view to see. We are usually unable to make do with seeing a single aspect because of the complexity of social systems. That is, we must see the multiple aspects multi-dimensionally, implying the simultaneous use of many evaluation items, each of which reflects some aspect of objects.

As a tool for multi-dimensionally comprehensive evaluation, we have been employing the weighted sum of evaluation items when each of which is quantitative. However, it is often not easy to define such *a priori* weighting. Particularly in social systems analysis, it is difficult to specify clear-cut weighting because of the complexity and variety of human preference. Therefore, we must say that social systems evaluation in terms of the weighted sum almost always has some arbitrariness however the weights are specified.

Many of problems for social systems evaluation seem to emerge from specifying weights as well as choosing evaluation items. Therefore, we would rather consider an alternative comprehensive evaluation without specifying the weights. This paper proposes such a new evaluation, *non-uniform evaluation*, by using a method called DEA (Data Envelopment Analysis).

*This is a revised version of the paper: Hashimoto, A., Non-uniform and comprehensive evaluation in social systems analysis. Inst. Policy & Plann. Sci. Discussion Pap. Ser. 779, Univ. Tsukuba, 1998.

†Phone/Fax: +81-298-53-5548; E-mail: hashimot@sk.tsukuba.ac.jp

2 Non-uniform evaluation

The new evaluation proposed here does not specify *a priori* weights, but employs flexibly defined weights corresponding to each object being evaluated. This is a relative evaluation in which each object can be evaluated in terms of the weights that would most advantage itself. It is also a non-uniform evaluation different from traditional ones, making a good use of each object's feature characteristics. That is, we can evaluate objects in terms of individual articles for sale, different by each object with varying characteristics. Today, in an age of individuality with diverse views of value, there seem to be many cases for which this non-uniform evaluation is preferable to the uniform one using *a priori* fixed weights with some arbitrariness.

Non-uniform evaluation can be considered in that we evaluate all objects in the same single frame, but use different measures, i.e., evaluation criteria, in evaluating individual objects. Evaluating objects requires value judgements based on some evaluation criterion. The criterion, of course, has to be fair, but would not have to be uniform across objects. Non-uniform evaluation certainly uses different criteria by object, but they can be considered to be equal in the sense of each object's best suiting criterion. Therefore, we think it fair and reasonable to employ non-uniform evaluation in social systems analysis.

We use *DEA* as a tool that realizes non-uniform evaluation. DEA is, by Cooper (1996), one of the DEA originators, a "data oriented approach" for evaluating the performance of a collection of entities called DMUs (Decision Making Units) which are regarded as responsible for converting inputs into outputs. From a methodological point of view, this paper claims that DEA could be applicable to fields other than the above-stated production efficiency. That is, focusing on DEA as a tool for more general evaluation in the next section, we consider non-uniform evaluation in social systems analysis through application cases illustrated in the fourth section.

3 DEA as a non-uniform evaluation tool

3.1 DEA basic model

DEA, developed first by Charnes *et al.* (1978), is a mathematical programming technique for measuring the relative *efficiency* of DMUs, i.e., objects, with multiple *inputs* and multiple *outputs* [see e.g., Boussofiane *et al.* (1991), Charnes *et al.* (1994) for overviews]. The basic model for computing efficiency score $h_{j_0}(0 < h_{j_0} \leq 1)$ to target DMU j_0 is formulated as the following fractional programming problem with decision variables u_r, v_i:

$$\text{Maximize} \quad h_{j_0} = \frac{\sum_{r=1}^{t} u_r y_{rj_0}}{\sum_{i=1}^{m} v_i x_{ij_0}} \tag{3.1a}$$

$$\text{subject to} \quad \frac{\sum_{r=1}^{t} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1,\ j = 1, ..., n, \tag{3.1b}$$

$$u_r, v_i > 0,\ r = 1, ..., t, i = 1, ..., m, \tag{3.1c}$$

where y_{rj} = the amount of output r from DMU j; x_{ij} = the amount of input i to DMU j; u_r = the weight given to output r; v_i = the weight given to input i; n = the number of DMUs; t = the number of outputs; m = the number of inputs.

DEA efficiency scores to all the DMUs can be found by solving problem (3.1) (we can solve this through converting into a linear programming problem) n times, setting each DMU as target DMU j_0 in turn. Here, DMUs j_0 with the maximum $h^*_{j_0} = 1$ are judged *DEA efficient*, while the other DMUs with $h^*_{j_0} < 1$ are *DEA inefficient*.

Model (3.1) has a structure in which the efficiency is defined as a ratio of the weighted sum of outputs to that of inputs, and the weights are variable. Target DMU j_0 being evaluated is measured whether its efficiency can rank top ($h^*_{j_0} = 1$) or how near it can come closer to the top when the DEA score to the top is supposed to be unity ($h^*_{j_0} < 1$). That is, each DMU is relatively evaluated in terms of the weights that best suit itself, and the number of DMUs judged DEA efficient is usually multiple.

3.2 DEA evaluation beyond efficiency analysis

DEA examines how efficiently DMUs convert multiple inputs into multiple outputs. That is, any DMU producing more outputs with fewer inputs is judged relatively efficient. However, DEA models do not necessarily assume such organic relationships between inputs and outputs as those in production [see model (3.1)]. Therefore, replacing inputs with *negative* evaluation items (the smaller the value, the better) and outputs with *positive* evaluation items (the greater the value, the better), yields a combined evaluation of these items. This is a non-uniform and comprehensive evaluation in terms of the weighted sum in that it replaces a uniform evaluation using an *a priori* weighting system with a flexibly defined weighting system corresponding to each DMU.

We should here note how the various perspectives in the current paper differ from that of production efficiency in the usual DEA sense. In production, an input is something going into the system while an output is something coming out of the system, and there exist organic relationships between inputs and outputs. On the other hand, we may choose evaluation items without considering such relationships between negative and positive evaluation items. There can be no outputs without inputs in production. Yet here, it is possible, for example, to have only negative (i.e., *pure input* DEA model) or only positive evaluation items (i.e., *pure output* DEA model).

This interpretation of DEA model accompanied by a DEA application case has been presented by the author and Ishikawa (Hashimoto and Ishikawa, 1993). Separately from us, Adolphson *et al.* (1991) have also stated such conception as a broader view of DEA, but they have gotten no further than suggesting the possibility of the pure-output and the pure-input DEA models. Therefore, the application cases of non-uniform evaluation shown in the next section are all ones carried out by those including the author. Besides those cases, we can find few "non-standard" applications beyond efficiency analysis though an increasing number of DEA applications have been reported. Prior to that, we explain two DEA extended models necessary for demonstrating our non-uniform evaluation cases.

3.3 DEA extended models

3.3.1 DEA/AR model

DEA is able to define a weighting system for inputs and outputs corresponding to a target DMU. This approach is in sharp contrast to the unified and uniform evaluation of using an *a priori* weghting system. A compromise between these two approaches is represented by *DEA/AR (DEA/ Assurance Region)* analysis (Thompson *et al.*, 1986). In model (3.1), we can discriminate the importance of evaluation items by bounding the ratios of weights, e.g., $u_1/u_2 \geq 2$. DEA/AR analysis aims at a more realistic analysis by incorporating experiences and expert opinions in the shape of constrained weight system. We should here note that the DEA/AR model requires some normalization of input and output data.

3.3.2 DEA exclusion model

It is a distinctive feature of DEA that multiple DMUs have the maximum DEA score unity. In *DEA exclusion* model (Andersen and Petersen, 1993), the DMU being evaluated is excluded from the comparison set by letting constraint (3.1b) be for $j = 1, ..., n, j \neq j_0$. Through this, for DMU j_0 with DEA score $h^*_{j_0} < 1$ whose comparison reference is the top DMU not j_0 itself, its DEA exclusion score is equal to $h^*_{j_0}$. While for DMU j_0 with $h^*_{j_0} = 1$, its comparison reference would be a DMU in the second place, and the DEA exclusion model measures how far the DMU j_0 can have a lead on the second place when the score of the second place is supposed to be unity (Hashimoto, 1997). Therefore, the DEA exclusion model allows DEA scores for DEA efficient DMUs to exceed unity, so that it can discriminate DEA efficient DMUs in terms of the DEA exclusion scores.

4 Some non-uniform evaluation cases

We can grasp an outline of a DEA analysis through knowing its DMUs, inputs and outputs. Therefore, in each of the following application cases, we first show those three though the cases do not necessarily have inputs and outputs in the production efficiency sense.

4.1 QOL analysis of Japan's prefectures (Hashimoto and Ishikawa, 1993)

DMUs: 47 prefectures of Japan

Inputs: 1 Suicide (Suicides per population)
2 Crime (Criminal cases recognized by police per population)
3 Traffic accident (Persons killed in road traffic accidents per population)
4 Bankruptcy (Bankruptcy cases per company)
Outputs: 1 Hospital bed (Hospital beds per population)
2 Income (Per capita prefectural income)
3 Water quality (Proportion of water resources achieving national standard)
4 House space (Per capita area of house)

This study relatively evaluates QOL (Quality-Of-Life) of Japan's 47 prefectures in terms of eight social indicators, which are chosen as those reflecting four aspects: health, safety, economy and environment. QOL should be measured through multiple aspects, so that we usually use the indicators' weighted sum as an integrated measure. But, it is difficult to define such *a priori* weighting because QOL is a personal, subjective and/or sensitive matter. Therefore, we can consider employing a DEA evaluation without using *a priori* fixed weights.

We apply negative and positive social indicators as inputs and outputs in DEA, respectively. Therefore, any prefecture with greater positive and smaller negative indicators than others is judged relatively livable (*DEA livable*). We should here note that we may choose negative and positive social indicators without considering organic relationships between them. Further, we need not always select the same number of negative and positive indicators as of this study, and we need not make the directions of negative and positive indicators coincide in case of integration. This study would have been the first DEA analysis in terms of multiple inputs and multiple outputs with no relations.

The results of the analysis found 26 DEA livable prefectures out of the 47, each of which has its own feature characteristics. Further, a DEA/AR analysis reduced the number of DEA livable prefectures into twelve. In this way, we can avoid indiscriminately uniform comparisons of prefectural QOL in terms of non-uniform evaluation.

4.2 Time series analysis of Japan's QOL (Hashimoto and Kodama, 1997)

DMUs: 35 years (1956-1990)
Inputs: 1 Suicide (Suicides per population)
2 Crime (Criminal cases recognized by police per population)
3 Traffic accident (Persons killed in road traffic accidents per population)
4 Unemployment (Ratio of totally unemployed persons to labor force)
Outputs: 1 Life expectancy (Life expectancy at birth)
2 Income (Per capita national income deflated by consumer price index)
3 Forest area (Per capita area of forest)
4 Water service (Diffusion rate of water service)

We here examine whether QOL of Japan has gotten better for the period 1956-1990 by observing its transition. In terms of time series data on social indicators for the period, we would like to relatively evaluate Japan's QOL like the analysis shown in subsection 4.1. A time series DEA analysis treating each year as a separate DMU makes it possible. For such DEA analyses with separate years as DMUs, we can find Cooper *et al.* (1995) as a DEA efficiency analysis.

The study found 20 DEA livable years out of the 35. Two DEA/AR analyses considering the trade-off between economy and environment designated eight years as the best-balanced years among the 20, most of which are in the period 1966-1980. It is concluded that we cannot simply say that QOL of Japan has gotten better or worse for 1956-1990.

4.3 Baseball batters evaluation (Hashimoto, 1993)

DMUs: 66 pro baseball batters
Inputs: 1 At bat (Number of being at bat)
2 Double play (Number of having double and triple plays)
Outputs: 1 Hit (Number of hits)
2 Walk (Number of walks)
3 Steal (Number of bases stolen)
4 Sacrifice (Number of sacrifices)

5 RBI (Number of runs batted in)

This study evaluates pro baseball batters. In assessing them, we tend to focus our viewpoints on *batting average*, *home runs* and *RBI (Runs Batted In)*. These three dimensions are traditionally used to demonstrate batting prowess as the *triple crown* frontier. However, considering that a batter's mission is contributing to runs, we should take also steals, sacrifices, etc. into account. When we see those contributions comprehensively, a uniform evaluation would not be fair because each batter's article for sale is different. Therefore, it is desirable to evaluate batters making a good use of their feature characteristics.

In this DEA analysis, we employ the basic consideration of how efficiently a batter converts the given opportunity of being at bat into contribution to runs. The contribution is further divided into *getting bases* (hit and walk), *advancing bases* (steal and sacrifice), *scoring* (RBI) and negative contribution (double play). Here, we consider double play as an input because the smaller the value, the better, making it apart from efficiency analysis. Home run is not employed as an output because we generally observe a very high correlation with RBI. These are the difference from Mazur (1994) DEA-evaluating batters in terms of the traditional triple crown frontier.

The results of the analysis found 16 *DEA outstanding* batters out of the 66, which include those far from the triple crown frontier as well as *sluggers*. A DEA/AR analysis discriminating the importance of outputs and considering at-bat as nondiscretionary selected three batters as DEA/AR outstanding.

4.4 DEA selective examination system (Hashimoto, 1996)

DMUs: 50 applicants
Inputs: None
Outputs: 1 Mathematics (Examination score on Mathematics)
2 Science (Examination score on Science)
3 Japanese (Examination score on Japanese)
4 Social studies (Examination score on Social studies)
5 English (Examination score on English)

We propose a selective examination system applicable to the university entrance examinations of Japan in the case where the following opinion is taken into consideration: University students are too uniform in these days, i.e., "unconventional" students are seldom seen; To make university students be rich in diversity and variety, we should admit also those who are brilliant in some academic subject but not necessarily in all as "masters of an art".

That is, the examination system here is required to be one that can select "masters of an art" as well as "jacks of all trades" (scoring quite well across all the subjects) as successful candidates. The examination is on paper and on five key academic subjects in Japan (see outputs). Further, the number of candidates taking the examination is 50, and the allocation of successful candidates is 25. Usual examination systems would select "jacks of all trades" so as to fill up the allocation in order of the total examination score. It could not select "masters of an art" because of their relatively low total-scores. A DEA examination system enables this.

The pure output DEA model applied here found three DEA efficient candidates out of the 50. They are *DEA brilliant* in the sense that each can rank top in terms of the weights optimal to himself/herself, i.e., no other candidates dominate them. Therefore, removing them as successful candidates of the first stage, we apply DEA to the remaining candidates. Repeating this selection up to the stage at which the cumulative number of successful candidates is equal to the allocation or more.

The results selected a total of 29 candidates up to the fourth selection stage (first: 3; second: 4; third: 10; and fourth: 12). These successful candidates certainly include "masters of an art", so that we can here see a property peculiar to DEA vs other comprehensive evaluation tools, i.e., multiple candidates with various featured characteristics can rank top. If we must adjust the number of successful candidates exactly to the allocation, we may leave out four amongst the candidates selected at the last stage using techniques for discriminating DEA efficient DMUs [e.g., Andersen and Petersen (1993), Cook *et al.* (1992)].

Rather, we should here note the fairness among the candidates. In this DEA examination system, every candidate can equally be evaluated in terms of criteria that best suit himself/herself. We think it fair though the evaluation criteria used are not uniform across the candidates.

4.5 Ranked voting systems analysis (Hashimoto, 1997)

DMUs: 14 candidates
Inputs: None
Outputs: 1 First place (Number of first place votes)
2 Second place (Number of second place votes)
3 Third place (Number of third place votes)
4 Fourth place (Number of fourth place votes)
5 Fifth place (Number of fifth place votes)

This study addresses ranked voting systems in which each voter selects and ranks the top t candidates. The problem is to determine an ordering of all n candidates by obtaining a total score $s_j = \sum_{r=1}^{t} u_r y_{rj}$ for each candidate $j, j = 1, ..., n$, where y_{rj} is the number of r-th place votes that candidate j receives, and $u_r, r = 1, ..., t$ is the sequence of weights given to the r-th place vote. Because of no established way to determine the weights, many arbitrary choices of the sequence of weights can exist. The well known Borda method, $u_r = t - r + 1$, is an example.

The pure output DEA model with the r-th place vote as the r-th output makes it possible not to specify such a sequence of weights with arbitrariness. A DEA model for ranked voting should originally be a DEA/AR model assuming at least $u_r > u_{r+1} > 0, r = 1, ..., t-1$. The DEA/AR model may have several candidates tied for the first place as DEA/AR efficient. To resolve the problem of ties, Cook and Kress (1990) introduced a function implying the minimum gap between successively ranked weights called the discrimination intensity function. But this time, the arbitrariness for specifying the function still remained.

The study (Hashimoto, 1997) shows that we can obtain a total ordering of candidates specifying nothing arbitrary by using a DEA/AR exclusion model. The results of the analysis using data quoted from Stein *et al.* (1994) certainly demonstrated that a total ordering of candidates different from the Borda one is obtained. There, it newly introduces a satisfactory interpretation of the criterion on which we discriminate among DEA/AR efficient candidates as how far the candidate being evaluated can have a lead on the candidate in the second place (see subsection 3.3.2). Further, it also states that not DEA exclusion but the DEA/AR exclusion model used in this study can resolve the shortcoming of the exclusion model that outlying DMUs are ranked too high.

5 Summary and conclusions

While the idea of comparing objects by taking a weighted sum of their attributes is commonplace, the idea that each object may have the freedom to choose its own optimal weights is not commonplace. This paper proposed a new comprehensive evaluation using DEA from the latter idea, non-uniform evaluation, and demonstrated its application cases in social systems analysis.

From a methodological viewpoint, this paper showed through the cases that DEA can be employed in more general evaluation beyond production efficiency. Although a very large number of DEA related works have been reported, there exist few "non-standard" applications other than efficiency analysis. Especially, we can find no other DEA applications with multiple inputs and multiple outputs that have no relations between themselves.

As seen in the demonstrated application cases, we had better apply non-uniform evaluation to the cases where objects being evaluated have different feature characteristics, i.e., different articles for sale, and/or established clear-cut weights are difficult to be specified. Such situations would rather be general recently, so that non-uniform evaluation would enlarge its appropriate application fields in social systems analysis.

References

Adolphson, D. L., Cornia, G. C. and Walters, L. C. (1991) A unified framework for classifying DEA models. In H. E. Bradley (ed.), *Operational Research '90* (pp. 647-657), Oxford: Pergamon Press.

Andersen, P. and Petersen, N. C. (1993) A procedure for ranking efficient units in data envelopment analysis. *Management Science* **39**, 1261-1264.

Boussofiane, A., Dyson, R. G. and Thanassoulis, E. (1991) Applied data envelopment analysis. *European Journal of Operational Research* **52**, 1-15.

Charnes, A., Cooper, W. W., Lewin, A. Y. and Seiford, L. M. (eds.) (1994) *Data Envelopment Analysis: Theory, Methodology, and Application.* Boston: Kluwer Academic Publishers.

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

Cook, W. D. and Kress, M. (1990) A data envelopment model for aggregating preference ranking. *Management Science* **36**, 1302-1310.

Cook, W. D., Kress, M. and Seiford, L. M. (1992) Prioritization models for frontier decision making units in DEA. *European Journal of Operational Research* **59**, 319-323.

Cooper, W. W. (1996) Data envelopment analysis. In S. I. Gass and C. M. Harris (eds.), *Encyclopedia of Operations Research and Management Science* (pp. 145-150), Boston: Kluwer Academic Publishers.

Cooper, W. W., Kumbhakar, S., Thrall, R. M. and Yu, X. (1995) DEA and stochastic frontier analyses of the 1978 Chinese economic reforms. *Socio-Economic Planning Sciences* **29**, 85-112.

Hashimoto, A. (1993) Evaluating baseball batters using DEA (in Japanese). *Communications of the Operations Research Society of Japan* **38**, 146-153.

Hashimoto, A. (1996) A DEA selection system for selective examinations. *Journal of the Operations Research Society of Japan* **39**, 475-485.

Hashimoto, A. (1997) A ranked voting system using a DEA/AR exclusion model: A note. *European Journal of Operational Research* **97**, 600-604.

Hashimoto, A. and Ishikawa, H. (1993) Using DEA to evaluate the state of society as measured by multiple social indicators. *Socio-Economic Planning Sciences* **27**, 257-268.

Hashimoto, A. and Kodama, M. (1997) Has livability of Japan gotten better for 1956-1990?: A DEA approach. *Social Indicators Research* **40**, 359-373.

Mazur, M. J. (1994) Evaluating the relative efficiency of baseball players. In A. Charnes *et al.* (eds.), *Data Envelopment Analysis: Theory, Methodology, and Application* (pp. 369-391), Boston: Kluwer Academic Publishers.

Stein, W. E., Mizzi, P. J. and Pfaffenberger, R. C. (1994) A stochastic dominance analysis of ranked voting systems with scoring. *European Journal of Operational Research* **74**, 78-85.

Thompson, R. G., Singleton, Jr., F. D., Thrall, R. M. and Smith, B. A. (1986) Comparative site evaluations for locating a high-energy physics lab in Texas. *Interfaces* **16**, 35-49.