International Institute for Applied Systems Analysis • A-2361 Laxenburg • Austria
Tel: +43 2236 807 • Fax: +43 2236 71313 • E-mail: info@iiasa.ac.at • Web: www.iiasa.ac.at

INTERIM REPORT IR-98-032/June

Value Efficiency Analysis of Academic Research

Pekka Korhonen
Risto Tainio
Jyrki Wallenius

Approved by
Gordon MacDonald (macdon@iiasa.ac.at)
Director, *IIASA*

Abstract

We propose a systematic approach to analyzing academic research performance at universities and research institutes. The analysis of research performance is based on a set of (abstract) criteria which are relevant from the decision maker's point of view. The scales for these criteria are defined by means of concrete indicators. All indicators are, however, not necessarily quantitative. Qualitative information is quantified using appropriate analytical tools. Once the criteria and indicators have been agreed upon and quantified, data on the research units is collected and a Value Efficiency Analysis is performed. The efficiency of research units is defined in the spirit of Data Envelopment Analysis (DEA), complemented with a decision maker's (Rector in the European university system) preference information. This information is obtained by asking the decision maker to locate a point on the efficient frontier having the most preferred input and output values. Our approach and the accompanying Decision Support System enables a university to allocate resources more efficiently for its research units. Using data from the Helsinki School of Economics, we describe how our approach can be used.

Keywords: Multicriteria Analysis, Data Envelopment Analysis, Academic Research, Education, Linear Programming

Acknowledgments

This research was supported in part by grants from the Academy of Finland and the Foundation of the Helsinki School of Economics.

About the Authors

Pekka Korhonen is Project Leader of the Decision Analysis and Support Project at IIASA, and also Professor of Statistics at the Department of Economics and Management Science, Helsinki School of Economics and Business Administration.
Risto Tainio and Jyrki Wallenius are with the Helsinki School of Economics and Business Administration.

Value Efficiency Analysis of Academic Research

Pekka Korhonen
Risto Taiio
Jyrki Wallenius

1. Introduction

It is paradoxical that in spite of the obvious importance of research in research oriented universities and research institutes around the world, it seems very difficult to '*scientifically'* evaluate research. The problem is universal and several universities have developed their own approaches to this problem. Two extreme approaches are the process-oriented and the results-oriented approach. In the process-oriented approach, the focus is on the research process. The approach is based on the premise that a high quality research-process produces high quality research results. In the results-oriented approach, various quantitative '*yardsticks*' are used to measure research output/input. The number of publications, citations, invited talks at conferences, etc., are examples of such yardsticks. It is vitally important, however, that the yardsticks have been properly defined and that the evaluation system is not too mechanical and incorporates the decision maker's (DM) preferences with regard to the importance of the yardsticks.

Previous research on the topic includes, among others, Moed, Burger, Frankfort, and Van Raan (1985), Wallmark and Sedig (1986), Wallmark, McQueen and Sedig (1988), Cave, Hanney, Kogan, and Trevett (1991), Önel and Saatcioglu (1995), Fandel and Gal (1997), and Lootsma and Bots (1997).

In the early 1990s the Ministry of Education in Finland signaled that in future years government research funding would to a larger extent be allocated to universities and schools demonstrating a track record of high quality research. This led the Research Development Group (TUTKE) at the Helsinki School of Economics, chaired by the Rector, to establish a two-person team[1] with the goal of developing an approach to evaluating research performance and to helping the administration of our school allocate research resources in the "best" possible way. Traditionally research (and teaching) resources at the Helsinki School of Economics have primarily been allocated according to the number of (research) positions in a unit, usually a (sub)department controlling its budget. Part of the funds have been allocated to projects, which have appeared

[1] The team comprised the first and the second author of this paper.

important and promising. Research performance at our school had previously not been systematically evaluated, although data on publications etc. had been collected. In the early days, the annual research catalogue containing such information, however, did not even classify different publications. In other words, refereed journal articles, interim reports and working papers, and articles in management oriented magazines received equal coverage in the research catalogue. The first task of the team was to develop a meaningful classification system for academic publications. This was followed by a thorough discussion of criteria and indicators, with which to measure research performance, and their relative importance. It was agreed upon that the evaluation would not take place at individual but research unit level.

In this paper we describe our results-oriented system for evaluating academic research. It consists of the following steps. (1) Definition of criteria and indicators which are used to measure research performance of research units. (2) Collection of appropriate data from the research units. If all data is not quantitative, we use existing analytical tools to quantify such qualitative data. (3) In the spirit of Data Envelopment Analysis (DEA), calculating Value Efficiency scores for each unit incorporating the DM's preferences regarding the importance of different outputs and inputs. Value Efficiency scores are based on identifying the DM's most preferred combination of outputs and inputs. Reflecting our own bias, we have used the Pareto Race/VIG system for this task. In a nutshell, Value Efficiency scores compare the inefficient units to units having the same value as the Most Preferred Solution. The resulting Value Efficiency scores are optimistic approximations of the true scores. For an explanation of the underlying procedure and the requisite theory, see Halme, Joro, Korhonen, Salo, and Wallenius (1998).

This paper consists of five sections. In Section 2 we develop the data base for the performance analysis. In Section 3 we review DEA and Value Efficiency Analysis. In Section 4 we discuss the research performance analysis at the Helsinki School of Economics,and in Section 5 conclusions are given. .

2. Development of the Data Base

2.1. Characterization of a 'Model' Research Unit

We begin with a concise characterization of a 'model' research unit, which can serve as an example to other units. In our concise characterization of a 'model' research unit, we use general terms so that a broad unanimity can easily be reached. The purpose of the characterization is to help identify relevant criteria.

A research unit whose members ***continuously produce high quality****, innovative and* ***internationally recognized*** *research, and who actively supervise* ***doctoral students*** *and actively take part in various activities of the* ***scientific community****.*

Based on our experience, it is not difficult to reach unanimity about such a characterization of a 'model' research unit, provided that we keep the discussion at a rather high level of abstraction as above. It is also important that we use a large enough set of attributes in the characterization. In our experience, it even seems possible to specify some jointly agreed upon aspiration levels for the attributes, if we do not engage in a discussion concerning the importance of achieving such aspiration levels. The aspiration levels are needed for making the discussion more concrete regarding abstract expressions such as: ‘high quality’ or ‘continuously produce’, etc.

At universities with several colleges or faculties, and research cultures, it may be better to provide more than one characterization of a ‘model’ research unit. At the Helsinki School of Economics, one characterization was accepted. When aspiration levels were specified for different attributes at our school, we have mentioned them in parenthesis. This more complete and concrete characterization of a ‘model’ research unit, accepted at our school is provided below.

The research problems investigated are relevant from the point of view of the school’s mission. The quality of research is internationally recognized, i.e. research results are accepted for publication in refereed journals (1-2 articles/year/researcher), and they have impact on the development of the research field (1-2 citations/year/researcher). Research is active (1-2 working papers/year/researcher) and interdisciplinary. Each professor has a number of active doctoral students (2-6) under his/her supervision, with whom joint publications are written. The research unit has established international research contacts, and well known foreign researchers visit the research unit, providing supervision and advice to doctoral students. The researchers are invited to foreign universities to give invited and key-note presentations in international seminars and conferences. The researchers enjoy a certain position in the scientific community. They are board members of professional associations, editors or members of editorial boards in professional journals and members of program and organizing committees of scientific conferences. Their expertise is solicited to review journal articles and the qualifications of colleagues applying for various scientific positions. The research results have real impact on teaching, keeping it timely and of high quality. Furthermore, private and public organizations have an interest in applying the research results in practice.

2.2 Evaluation Criteria and Indicators

We next define a set of **criteria**, which are sufficient to characterize the ‘model’ research unit. The criteria should be relevant to the DM. They should emphasize different aspects of research performance in the same spirit as our (abstract) characterization of the ‘model’ research unit did. In our characterization we have highlighted the keywords, which we will use as a basis for introducing the criteria. After several discussion sessions at the Helsinki School of Economics we ended up with the following set of criteria:

- Quality of Research
- Research Activity
- Impact of Research
- Activity in Educating Young Scientists (especially doctoral students)
- Activity in Scientific Community

It is interesting to note that our criteria are rather close to the criteria applied in the Quality Assessment of Academic Research in the Netherlands (Economics, 1995; Mathematics and Computer Science, 1997): Scientific Quality; Scientific Productivity; Relevance; and Long-term Viability.

It is important to use multiple criteria in the evaluation, because it is extremely difficult - if not impossible - to find a way to aggregate the criteria into one criterion. In addition, when using several criteria, we can take into account that the research units may be good in different ways, which depends on their development phase. For instance, if the research unit is new, it can be regarded as a high performer, if its general activity is high. On the other hand, an established unit may be considered good if the quality of research is stable, and it is active in supervising doctoral students. The general activity of such a unit may be low, because its members have many duties in the scientific community.

The above listed criteria are suitable for a general discussion, but too abstract for enabling the DM to make a concrete evaluation based on them. That’s why it is necessary to introduce concrete **indicators** (attributes, signals), which can be employed to make use of the criteria more systematic. Indicators are more concrete than criteria in the sense that we can somehow more or less objectively ‘measure’ alternatives with them. An indicator is also a ‘*sign*’ that a unit is performing well in terms of a criterion. Furthermore, the set of indicators associated with a criterion can also be used to introduce a scale for the underlying criterion. Note that a criterion implies an indicator, but not necessarily the converse. It is important that the criteria are complete, nonredundant, and that their number is minimal to the extent possible.

It is important that the indicators contain enough (objective) information about the values of the criteria, but even more important is that the indicators *cannot be manipulated* without an influence on the performance of the unit in terms of the corresponding criterion (Fig. 1). For instance, the number of visitors in a research unit is not a proper indicator for criterion ‘Quality of Research’, because the research unit can easily manipulate it without it having any influence on the criterion. The number of papers in refereed journals partly suffers from the same problem. If it is used as an indicator, the researchers may try to do their best by having their papers published in any refereed journal, without considering the quality of the journal. The number of citations may also be problematic, because sometimes "friends" cite each others' papers. A (small) number of citations does not necessarily imply high quality of research, although it implies that you are (somewhat) known in the scientific community. Our discussion shows that it is difficult to find indicators which are perfect. As a general rule, we propose the use of indicators, which cannot be manipulated without making at least some contribution to the underlying criterion/criteria. From the above mentioned set, we should definitely drop the "Number of Visitors" as an indicator of the quality of research.

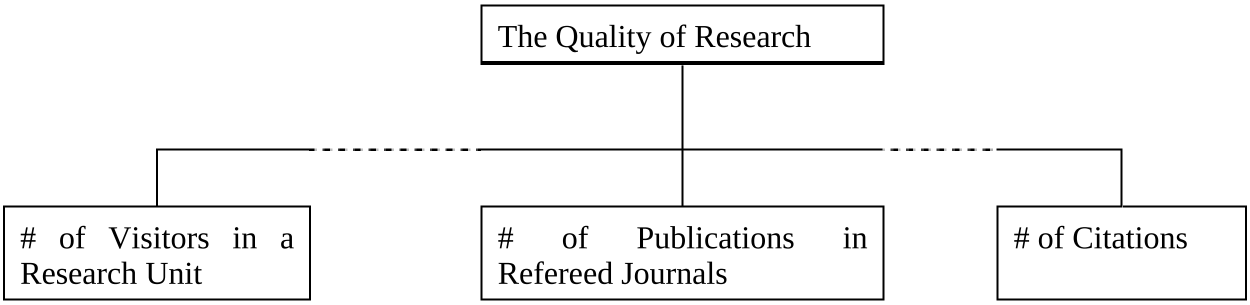

Figure 1: An Example of the Use of Indicators

After careful deliberations, we propose the use of the following indicators. Note that the same indicator can be relevant for several criteria.

a) Criterion: Quality of Research

articles in international refereed journals

scientific books and chapters in scientific books published by internationally well known publishers

citations

b) Criterion: Research Activity

publications exceeding a minimum quality standard (articles in refereed journals and scientific books and chapters in books)

papers in conference proceedings, national reports, reports in nonrefereed national journals, working papers and other unpublished reports

conference presentations

c) Criterion: Impact of Research

citations by other researchers (in journal articles, books, published conference proceedings, and Ph.D. dissertations)

invited and plenary presentations at international conferences

number of foreign co-authors in journal articles

d) Criterion: Activity in Educating Young Scientists

doctoral degrees produced

number of doctoral students supervised

e) Criterion: Activity in Scientific Community (not currently used)

memberships in editorial boards

edited books and special issues of journals

service as an expert

organizing scientific conferences, membership of program committees etc.

The above hierarchy provides a basis for a systematic evaluation of research, but also for a structured discussion. The literature does not often make a difference between criteria and indicators. For instance,research funds that a unit has received are often used as a performance criterion for research itself. Is the purpose of research to maximize the research funds of a department? Perhaps research funding is rather an indicator for ‘Impact of Research’, ‘Relevance of Research’, or something like that. It is important first to think what the relevant criteria are and which indicators are needed in their evaluation.

When it is desired to carry out a systematic and quantitative evaluation, we have first to introduce the scales for the criteria and locate alternatives using these scales. In case all indicators are quantitative, our problem is to define a function which aggregates the values of the indicators into a criterion scale. If some indicators are qualitative, we first have to quantify them by using appropriate tools. One of the simplest ways to aggregate the indicator values is to use weighted sums. It is a commonly used method, but there are many problems in its use. How to scale the values of the indicators? Do the weights describe what we or the DM think they do? How to take into account the dependence of criteria? At the Helsinki School of Economics, we used the weighted sums, but tried to take into account the problems associated with their use to the extent possible.

2.3 The Data

In total, 18 research units at the Helsinki School of Economics were included in the study. Many of the units represented functional business school areas, such as Organization & Management, Accounting, Finance, Marketing, Logistics, etc. Some of the units were interdisciplinary by nature, such as Management Science and Quantitative Methods. One of those 18 units is a unit called Basic Research Institute. It consisted of 1-2 professors and a number of graduate students from major areas represented at the school. It is specialized only for research. That's why it is clearly an outlier among the units representing traditional business school areas.

For reasons of confidentiality, the identity of the research units has been disguised. For the same reason, we do not publish the original values of the indicators. In the performance evaluation, we used four criteria: Quality of Research, Research Activity, Impact of Research, and Activity in Educating Doctoral Students. We were not able to obtain reliable information about the indicators comprising the fifth criterion, and that is why it is ignored in the analysis. To introduce the scales for the criteria, we used the indicators mentioned in subsection 2.2.

First, the values of all indicators were scaled into the [0, 1] range, so that the best value corresponded to one. Table 1 provides a summary of the data, which is used as the basis of the analysis. The information has been collected from different data files. Next we explain how the scores for the criteria in Table 1 have been calculated. The nine members of TUTKE at the Helsinki School of Economics were then asked to evaluate the relative importance of various indicators with regard to each criterion. (They were informed that the best value of each indicator was always the same.) Both direct weighting and the Analytic Hierarchy Process (AHP) were used (Saaty, 1980). The average of the weights of the nine members generated using the AHP were taken as the final weights. The weighted sum of the indicators was used as a scale for each criterion.The last column in Table 1 is an input measure. It is the estimated monthly cost in Finnish Marks of producing the research output. We can see that the size of the units varies considerably.

Table 1: Criterion Values as the Weighted Sums of Indicators and Resources

	Ref. Art.	**Books**	**Citat**	**C_1**	**Art + Books**	**Other**	**Conf Pres**	**C_2**	**Citat**	**Invit. Pres**	**For. Co-A**	**C_3**	**Dis.**	**Sup.**	**C_4**	**1000 Fmk/ Month**
Weights:	**0.55**	**0.27**	**0.18**		**0.60**	**0.23**	**0.17**		**0.48**	**0.27**	**0.25**		**0.65**	**0.35**		
A	0.70	1.00	0.10	**67**	1.00	1.00	1.00	**100**	0.10	1.00	0.67	**48**	1.00	1.00	**100**	70
B	0.67	0.03	0.05	**38**	0.36	0.38	0.33	**36**	0.05	0.50	0.67	**32**	0.30	0.15	**25**	32
C	0.03	0.13	0.00	**5**	0.10	0.08	0.00	**8**	0.00	0.00	0.00	**0**	0.13	0.00	**9**	34
D	0.30	0.16	0.01	**21**	0.25	0.26	0.12	**23**	0.01	0.05	0.00	**2**	0.52	0.04	**35**	101
E	0.90	0.13	0.06	**54**	0.54	0.13	0.08	**37**	0.06	0.00	0.25	**9**	0.04	0.08	**6**	25
F	0.20	0.95	0.38	**43**	0.71	0.77	0.23	**64**	0.38	0.23	0.33	**33**	0.70	0.27	**55**	64
G	0.50	0.24	0.06	**35**	0.41	0.50	0.35	**42**	0.06	0.41	0.17	**18**	0.83	0.31	**65**	46
H	0.00	0.05	0.02	**2**	0.03	0.14	0.00	**5**	0.02	0.00	0.00	**1**	0.30	0.08	**23**	25
I	0.30	0.24	0.24	**27**	0.31	0.47	0.40	**36**	0.24	0.14	0.17	**19**	0.26	0.23	**25**	28
J	0.47	0.32	0.00	**34**	0.44	0.41	0.00	**36**	0.00	0.00	0.00	**0**	0.17	0.00	**11**	23
K	0.07	0.13	0.00	**7**	0.12	0.03	0.00	**8**	0.00	0.00	0.00	**0**	0.09	0.04	**7**	7
L	0.40	0.45	0.42	**42**	0.49	0.30	0.50	**45**	0.42	0.41	0.17	**35**	0.70	0.15	**51**	68
M	0.00	0.16	0.01	**4**	0.10	0.02	0.02	**7**	0.01	0.05	0.00	**2**	0.00	0.00	**0**	8
N	0.27	0.05	0.13	**18**	0.17	0.02	0.23	**15**	0.13	0.45	0.33	**27**	0.00	0.08	**3**	15
O	0.43	0.00	0.08	**25**	0.22	0.17	0.23	**21**	0.08	0.45	0.33	**25**	0.30	0.15	**25**	37
P	0.90	0.26	0.92	**74**	0.63	0.18	0.33	**47**	0.92	0.55	0.92	**82**	0.00	0.15	**5**	29
Q	0.13	0.74	0.01	**27**	0.54	0.37	0.62	**51**	0.01	0.73	0.00	**20**	0.00	0.00	**0**	12
R	1.00	0.11	1.00	**76**	0.58	0.58	0.44	**55**	1.00	0.05	1.00	**74**	0.00	0.08	**3**	119
			Σ =	**602**				**636**				**427**			**447**	

When we use the weighted sums of indicators to find the scores for criteria, we have implicitly assumed that the units are homogeneous enough. Unfortunately, this is not the case at universities. We have assumed that it is possible to compare each research unit to the best performing unit with regard to the indicator in question. If, e.g., the unit, Law, has published only 5 articles a year in international journals and the best performing research unit, Marketing, has published 50 articles a year, the performance of the unit, Law, Faculty is only 10% from its maximal performance with respect to this specific indicator. However, Law and Marketing are very different scientific disciplines, and therefore it is not fair to require that they have the same values for different indicators. To correct this bias, we could for instance use adjusted indicator values, where each value is adjusted by comparing it to the best value of its own discipline (in other schools).

3. The Method

3.1 Data Envelopment Analysis

Assume we have n decision making units (DMU) each consuming m inputs and producing p outputs. Let $\mathbf{X} \in \Re_{+}^{m \times n}$ and $\mathbf{Y} \in \Re_{+}^{p \times n}$ be the matrices, consisting of nonnegative elements, containing the observed input and output measures for the DMUs. We further assume that there are no duplicated units in the data set. We denote by $\boldsymbol{x}_j$ (the jth column of $\mathbf{X}$) the vector of inputs consumed by DMUj, and by x_{ij} the quantity of input i consumed by DMUj. A similar notation is used for outputs. Furthermore, we denote $\boldsymbol{1} = [1, ..., 1]^{\mathrm{T}}$.

The traditional CCR-models, as introduced by Charnes et al. [1978] are fractional linear programs which can easily be formulated and solved as linear programs. Later Banker, Charnes and Cooper [1984] developed the so-called BCC models with variable returns to scale. The CCR and BCC models are the basic model types in DEA. In this paper, we consider solely output oriented BCC models. In BCC-models, the efficiency of a DMU is determined by maximizing outputs subject to given input levels. The discussion, with appropriate modifications, holds for output-oriented CCR-models, input oriented CCR- and BCC-models, and other DEA-models as well. The output oriented BCC-models are given in (3.1a) and (3.1b). Note that following Charnes and Cooper, the original primal formulation is called the dual and vice versa.

Output-Oriented BCC Primal **($\mathrm{BCC_P}$ - O)**	**Output-Oriented BCC Dual** **($\mathrm{BCC_D}$ - O)**
$\max Z_O = \theta + \varepsilon(\boldsymbol{1}^T\boldsymbol{s}^+ + \boldsymbol{1}^T\boldsymbol{s}^-)$ [2)] s.t. (3.1a) $\mathbf{Y}\lambda - \theta\boldsymbol{y}_0 - \boldsymbol{s}^+ = \boldsymbol{0}$ $\mathbf{X}\lambda + \boldsymbol{s}^- = \boldsymbol{x}_0$ $\boldsymbol{1}^T\lambda \le 1$ $\lambda, \boldsymbol{s}^-, \boldsymbol{s}^+ \ge \boldsymbol{0}$ $\varepsilon > 0$ ("Non-Archimedean") [3)]	$\min W_O = \nu^T\boldsymbol{x}_0 + u$ s.t. (3.1b) $\mu^T\boldsymbol{y}_0 = 1$ $-\mu^T\mathbf{Y} + \nu^T\mathbf{X} + u\boldsymbol{1}^T \ge \boldsymbol{0}^T$ $\mu, \nu \ge \varepsilon\boldsymbol{1}$ $\varepsilon > 0$

A DMU is efficient iff $Z^* = 1$ and all slack variables $\boldsymbol{s}^-$, $\boldsymbol{s}^+$ equal zero; otherwise it is inefficient (Charnes et al. 1994).

---

[2] For clarity, throughout the paper we assume that the units of all slacks are the same. See Thrall (1996) for a discussion.

[3] For more details, see Arnold, Bardhan, Cooper, and Gallegos [1997].

3.2 Value Efficiency Analysis

The idea of Value Efficiency Analysis is to incorporate the DM's preference information regarding a desirable combination of inputs and outputs into the analysis. This is in contrast with traditional DEA, which assumes that no output or input is more important than another. As explained in Halme et al. (1998), the preference information is incorporated via the Most Preferred Solution, i.e. a (virtual or existing) DMU on the efficient frontier having the most desirable values of inputs and outputs. Theoretically the DM is assumed to have a (unknown) pseudoconcave value function $v(\boldsymbol{u})$, $\boldsymbol{u} = \begin{bmatrix} \boldsymbol{y} \\ -\boldsymbol{x} \end{bmatrix} \in \Re^{m+p}$, which is strictly increasing (i.e. strictly increasing in $\boldsymbol{y}$ and strictly decreasing in $\boldsymbol{x}$) and with a (local) maximal value $v(\boldsymbol{u}^*)$, $\boldsymbol{u}^* = \begin{bmatrix} \boldsymbol{y}^* \\ -\boldsymbol{x}^* \end{bmatrix} \in$ T, at the Most Preferred Solution $\boldsymbol{u}^*$, where T stands for the feasible set.

The purpose of Value Efficiency Analysis is to evaluate efficiency of each unit in relation to the indifference contour of that (unknown) value function passing through the Most Preferred Solution. The evaluation could be done easily, if we explicitly knew the DM's value function. However, generally in practice it is not realistic to assume that the value function is known or that it could reliably be estimated. That is why we approximate that indifference contour by using all possible tangent hyperplanes. Those hyperplanes define a new 'Efficiency Frontier' and in relation to this frontier efficiency is then defined using a standard DEA-technique. Mathematically this reduces to a straightforward application of linear programming. The resulting scores are called Value Efficiency Scores. Because of using an approximation described above for the indifference contour of the value function, the resulting Value Efficiency scores are always optimistic approximations of the true scores.

The basic idea of Value Efficiency Analysis is illustrated in Fig.2. We have five units (A, B, C, D, E), which produce two outputs and use the same amount of one input. In Fig. 2, the problem has been described in the output space. The efficiency measure in standard DEA is the ratio: $\frac{\mathrm{OB}}{\mathrm{OB}^1}$. We would like to evaluate the ratio: $\frac{\mathrm{OB}}{\mathrm{OB}^4}$, but because the value function is unknown, we are not able to do it. If we could approximate the indifference contour by a tangent, then we could use the ratio: $\frac{\mathrm{OB}}{\mathrm{OB}^3}$. Because we do not assume that it is possible in practice, we have to consider all possible tangents of the contour. This leads to the use of the ratio: $\frac{\mathrm{OB}}{\mathrm{OB}^2}$ as an approximation to the (true) Value Efficiency Score. Because this approximation is the best we can get, we will call this score simply *Value Efficiency Score*.

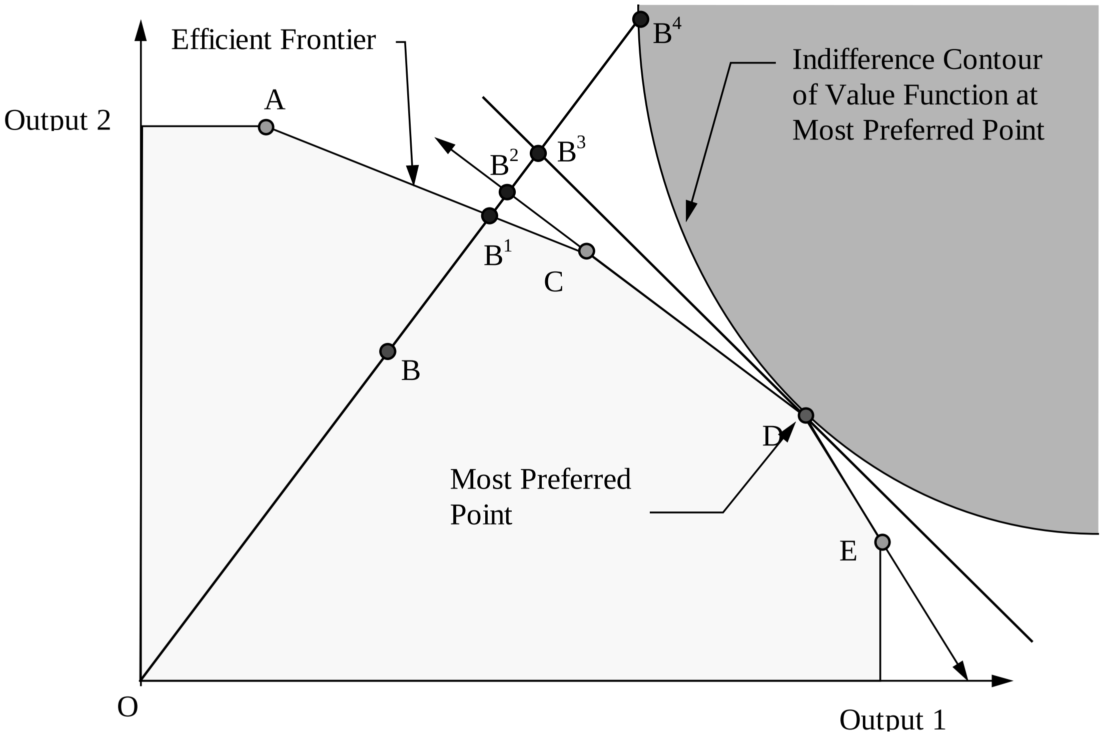

Figure 2: Illustration of Value Efficiency Analysis

Value Efficiency Analysis can be carried out as easily as a standard DEA using linear programming. A DMU is inefficient with respect to any strictly increasing pseudoconcave value function $v(\boldsymbol{u})$, $\boldsymbol{u} = \begin{bmatrix} \boldsymbol{y}^0 \\ -\boldsymbol{x}^0 \end{bmatrix}$ with a maximum at point $\boldsymbol{u}^*$, if the optimum value Z* of the following problem is greater than one:

$$
\begin{aligned}
&\max \quad Z = \sigma + \varepsilon(\mathbf{1}^T\boldsymbol{s}^+ + \mathbf{1}^T\boldsymbol{s}^-) \\
&\text{s.t.} \\
&\qquad \mathbf{Y}\lambda - \sigma\boldsymbol{y}^0 - \boldsymbol{s}^+ = \mathbf{0}, \\
&\qquad \mathbf{X}\lambda \quad + \boldsymbol{s}^- = \boldsymbol{x}^0, \\
&\qquad \mathbf{1}'\lambda + \mu \le 1, \\
(3.2) \quad & \\
&\qquad \boldsymbol{s}^-, \boldsymbol{s}^+ \ge \mathbf{0}, \\
&\qquad \varepsilon > 0, \quad (\text{“Non-Archimedean”}) \\
&\qquad \lambda_j \ge 0, \text{ if } \lambda_j^* = 0, \; j = 1,2, \ldots, n \\
&\qquad \mu \ge 0, \text{ if } \mu^* = 0
\end{aligned}
$$

where $\lambda^* \in \Lambda$, μ^* correspond to the Most Preferred Solution:

$$
\begin{aligned}
\boldsymbol{y}^* &= \mathbf{Y}\lambda^* \\
\boldsymbol{x}^* &= \mathbf{X}\lambda^*.
\end{aligned}
$$

Note that model (3.2) is the same as model (3.1a) with one exception. Nonnegativity constraints referring to the basic λ- and μ-variables with positive values corresponding to the Most Preferred Solution are relaxed. It is important that the Most Preferred Solution really lies on the efficient frontier. Otherwise the solution is unbounded (for a more detailed discussion, see Halme et al., 1998).

4. Analysis of Research Performance at the Helsinki School of Economics

We performed an efficiency analysis of 18 research units, using 4 output measures and one input measure as described in Table 1 (cols. C_1, C_2, C_3, C_4, and the last column). The data represent actual data from 1996. We first performed a standard output oriented BCC Data Envelopment Analysis (Banker, Charnes, Cooper, 1984) (Table 2). Four units (A, P, Q, and R) received the highest possible BCC efficiency score = 1. Table 2 also describes the reference set for each unit and the corresponding weights.

Table 2: Value Efficiency Analysis with an Output-Oriented Model

	BCC-Efficiency					BCC-Value Efficiency				
Depts.	efficiency	A	P	Q	R	efficiency	A	P	Q	R
A	1.00	1.00				1.00	1.00			
B	0.79	0.30	0.39			0.55	0.43	0.73		-0.16
C	0.17	0.49				0.08	1.04	0.38		-0.42
D	0.35	1.00				0.32	1.12			-0.12
E	0.88	0.02	0.81	0.03		0.75	0.05	1.01		-0.07
F	0.68	0.89	0.05	0.06		0.65	1.02	0.32		-0.34
G	0.98	0.66				0.54	1.19	0.16		-0.35
H	0.64	0.35				0.06	3.76	4.73		-7.49
I	0.77	0.32	0.13	0.17		0.41	0.85	0.56		-0.40
J	0.77	0.13	0.37	0.32		0.49	0.53	0.77		-0.30
K	0.86	0.08	0.04			0.11	0.63	0.91		-0.54
L	0.64	0.79	0.21			0.62	0.81	0.50		-0.31
M	0.23	0.00	0.07	0.52		0.07	1.12	0.63		-0.75
N	0.67	0.01	0.48			0.30	0.50	3.61		-3.11
O	0.58	0.42	0.27			0.38	0.64	0.95		-0.59
P	1.00		1.00			1.00		1.00		
Q	1.00			1.00		0.45	1.49	1.06		-1.55
R	1.00				1.00	1.00				1.00

We also performed a corresponding Value Efficiency Analysis of the units. For the analysis, Pareto Race (Korhonen and Wallenius, 1988) was used to freely search the efficient frontier corresponding to the BCC-model above. For using Pareto Race, we have first to formulate a multiple objective linear programming model which characterizes the efficient frontier of the BCC-model. The model is simple. We maximize all output-variables and minimize the input variable. The model is shown in Table 3.

Table 3: A Multiple Objective Linear Programming Model for Finding the Most Preferred Solution

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R		
Quality	67	38	5	21	54	43	35	2	27	34	7	42	4	18	25	74	27	76	max	51
Activity	100	36	8	23	37	64	42	5	36	36	8	45	7	15	21	47	51	55	max	67
Impact	48	32	0	2	9	33	18	1	19	0	0	35	2	27	25	82	20	74	max	55
Post-Grad	100	25	9	35	6	55	65	23	25	11	7	51	0	3	25	5	0	3	max	67
1000 Fmk	70	32	34	101	25	64	46	25	28	23	7	68	8	15	37	29	12	119	min	79
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	≤	1

Figure 2 shows the Pareto Race interface and the final solution at which the search was terminated. This point, which is a convex combination of units A and R, was taken as the Most Preferred solution. The corresponding values of the basic variables are A (0.748) and R (0.252). Both values are positive, and thus the nonnegativity constraints corresponding to these variables are relaxed in the Value Efficiency Analysis. The results of the Value Efficiency Analysis are given in Table 2.

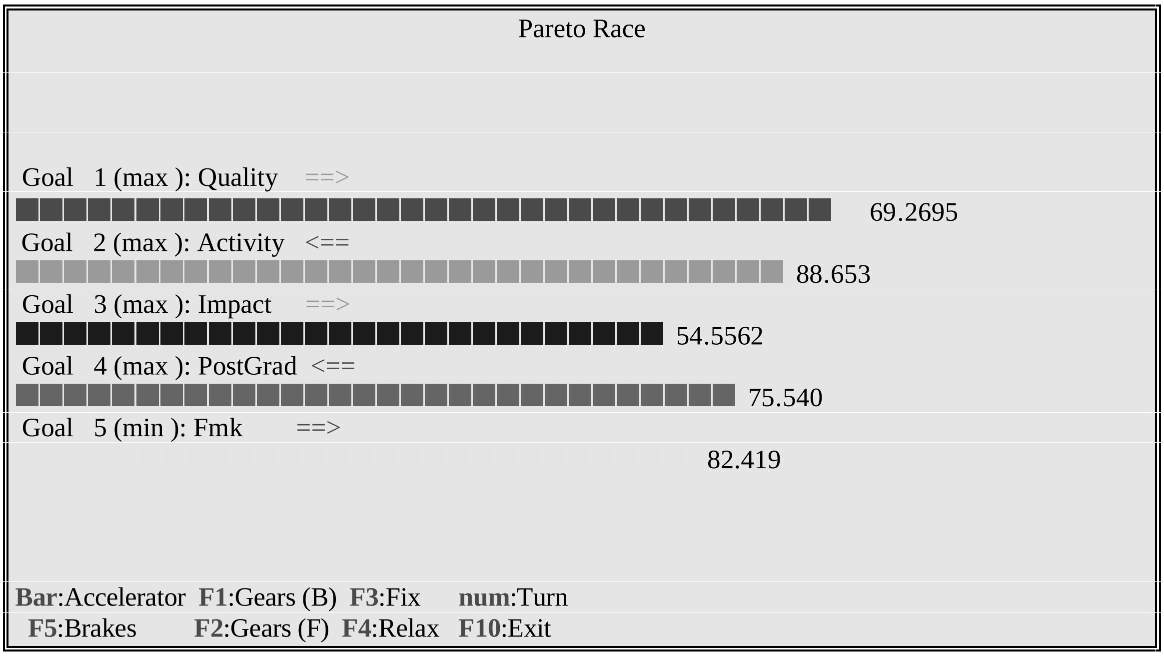

Figure 2: Pareto Race Screen

Only 3 of the four previously BCC efficient units remain Value Efficient. Commonly not all DEA efficient units remain Value Efficient. We also calculated the reference sets for the inefficient units as we did for the BCC-analysis. Note that all weights corresponding to unit R are negative. Usually only part of the weights of the units from which a nonnegativity constraint was relaxed are negative. It is also worth noting that all Value Efficiency Scores are not worse than the corresponding scores of standard DEA. This is intuitively understandable, because the 'Value Efficient Frontier' in a sense 'covers' the standard efficient frontier. Note that Value Efficiency Analysis has to be performed with the same model as standard efficiency analysis. One cannot use a

BCC-model to determine the most preferred solution and then use the CCR-model for Value Efficiency Analysis. In this specific case the solution is generally unbounded.

The data for 1997 are currently being updated so that our system can be used by the school's administration in allocating incentive money based on performance to the individual research units.

5. Conclusion

We have described a system to evaluate academic research performance. The problem is important, but difficult. Our approach is based on identifying a set of concrete indicators and aggregating them into decision-relevant criteria using importance weights. The output/input data is then analyzed using the concept of Value Efficiency, a novel procedure for incorporating the DM's preferences into Data Envelopment Analysis via the Most Preferred Solution. We have used real data from the Helsinki School of Economics to illustrate our ideas.

It may be of interest for European business schools to develop common standards for evaluating academic research. As future research we propose to find a set of criteria relevant for all universities, to find a set of indicators for each criterion, and to develop a system which makes it possible to standardize the 'yardsticks' at the European level within the disciplines and not within universities.

References

Arnold, V., Bardhan, I. , Cooper, W.W., and Gallegos, A. (1997), "Primal and Dual Optimality in Computer Codes Using Two-Stage Solution Procedures in DEA", Forthcoming in Aronson, J. and S. Zionts (Eds.): Operations Research: Models, Methods and Applications, Kluwer, Norwell. (A Volume in Honor of G.L. Thompson.)

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, 1078-1092.

Charnes, A., Cooper, W.W., and E. Rhodes (1978), "Measuring Efficiency of Decision Making Units", *European Journal of Operational Research* 2, 429-444.

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

Cave, M., Hanney, S., Kogan, M., and Trevett, G. (1991), *The Use of Performance Indicators in Higher Education,* Jessica Kingsley Publishers, London.

Fandel, G. and Gal, T. (1997), "Umvertailung der Mittel fur Lehre und Forschung unter den Universitäten", Diskussionsbeitrag 248, FB Wirtschaftswissenschaft, Fernuniversität Hagen, D-58084 Hagen, Deutschland.

Halme, M., Joro, T., Korhonen, P., Salo, S. and Wallenius, J. (1998), "A Value Efficiency Approach to Incorporating Preference Information in Data Envelopment Analysis", forthcoming in *Management Science*.

Korhonen, P. and Wallenius, J. (1988), "A Pareto Race", *Naval Research Logistics* 35, 615-623.

Lootsma, F. and Bots, P. (1997), "The Assignment of Scores for Output-Based Research Funding", Delft University of Technology, unpublished manuscript.

Moed, H. F., Burger, W., Frankfort, J., and Van Raan, A. (1985), "The Use of Bibliometric Data for the Measurement of University Performance", *Research Policy* 14, 131-149.

Saaty, T. (1980), *The Analytic Hierarchy Process*, McGraw Hill, New York.

Thrall, R.M. (1996), "Duality, Classification and Slacks in DEA", *Annals of Operations Research* 66, 109-138.

Wallmark, J. T. and Sedig, K. G. (1986), "Quality of Research Measured by Citation Method and by Peer Review -- A Comparison", *IEEE Transactions on Engineering Management 33*, 218-222.

Wallmark, J. T., McQueen, D. H. and Sedig, K. G. (1988), "Measurement of Output from University Research", *IEEE Transactions on Engineering Management* 35, 175-180.

Önel, A. and Saatcioglu, Ö. (1995), "Trend Analysis and Regression Modelling for the Assessment of Research Performance at a Turkish University", *Transactions on Operational Research* 7, 1-12.

Quality Assessment of Research, Economics. Association of the Universities in the Netherlands, Leidseveer 35, PO Box 19270, 3501 DG Utrecht, December 1995.

Quality Assessment of Research, Mathematics and Computer Science. Association of the Universities in the Netherlands, Leidseveer 35, PO Box 19270, 3501 DG Utrecht, September 1997.