Internet Based Benchmarking.

Peter Bogetoft
and
Kurt Nielsen

Internet Based Benchmarking

Peter Bogetoft and Kurt Nielsen
Working Paper

Department of Economics
The Royal Veterinary and Agricultural University
Denmark
April 2002

Abstract

We discuss the design of interactive, internet based benchmarking using parametric (statistical) as well as non-parametric (DEA) models. The user receives benchmarks and improvement potentials. The user is also given the possibility to search different efficiency frontiers and hereby to explore alternative improvement strategies. An implementation of both a parametric and a non parametric model are presented.

1 Introduction

Theorists and practitioners alike have devoted a lot of interest to benchmarking and relative performance evaluations in recent years. Theoretical advances, most notably the development of Data Envelopment Analysis (DEA) have gone hand in hand with new applications within all areas of society. On DEA alone, a 1999 bibliography lists more than 1000 studies[1], most of them published in good quality scientific journals. This is indeed an impressing record since DEA was only conceptualized some 24 years ago.

In general terms, benchmarking is the process of comparing the performance/activities of one unit against that of "best practice" units. DEA and other frontier evaluation techniques like Stochastic Frontier Analysis (SFA) are explorative data analysis and relative performance evaluation techniques that support advanced benchmarking. DEA was originally proposed by Charnes, Cooper and Rhodes(1978,79) and has subsequently been refined and applied in a rapidly increasing number of papers. For recent text books covering also related techniques like SFA, see Charnes, Cooper, Lewin and Seiford(1994), and Coelli, Rao and Battese(1998). Instead of using a traditional single dimensional performance indicator, say profit maximization, the DEA approaches use a multiple dimensional perspective and allow for different (efficient) combinations of products and services to be equally attractive. Also, instead of benchmarking against a theoretical engineering or a statistical average performance, DEA invokes a minimum of a priori assumptions and evaluates the performance of an unit to that of one or a combination of a few other, actual units. For these reasons, DEA has become a popular benchmarking approach when there is considerable uncertainty about the possibilities, e.g. the production structure of a hospital or university, and when the precise trade-off between different products and services, e.g. heart versus lung surgery, are hard to define.

There are multiple *uses* of relative performance evaluation and benchmarking. At least three distinct applications can be identified. Benchmarking can be used to get *general insight*, e.g. about the productive development of a given sector. For an extensive survey see Emrouznejad (2001). It can also be used to facilitate *decision making*, e.g. about the allocation of budgets within an organization Korhonen et.al. (2001) or the choice of consumption goods (or even communities) with respect to price and quality. There are a lot of papers in this area as well see e.g. Korhonen et al. (1992) for a review of Multi Criteria Decision Making (MCDM) based decision support systems.

---

[1]See e.g. www.deazone.com.

Thirdly, benchmarking and relative performance evaluation is the backbone of *incentive provision* in multiple agents contexts. For instance a regulator can use benchmarking to set prices or revenue caps (e.g. Bogetoft, 1994 and 2000).

Despite of the obvious success of the modern benchmarking approaches, we suggest that the potential of these techniques have not yet been fully realized. The reason is that most analyses still introduce a series of restrictions that the users may not accept or that they may want to modify over time and as the application of the analyses changes. It is therefore advantageous if the benchmarking can be tailored towards the specific circumstances and user.

To tailor the analysis to the specific user and context, a traditional report is insufficient. We need easy interaction with the different users. We therefore suggest to offer a *benchmarking environment* rather than a benchmarking report. To benchmark, the user interacts directly with a computer program. Moreover, to ease the communication of the benchmarking throughout an organization, an industry etc. the program shall be executed over the *internet*.

Of course, we cannot expect the traditional user to be trained in the pros and cons of different benchmarking approaches. An important challenge is therefore to allow flexibility via the benchmarking environment while at the same time offering more structure and guidance that the existing computer codes to support DEA, SFA and similar techniques.

The primary flexibility concerns the *benchmark* or *reference selection*. A traditional benchmarking exercise involves the selection of a reference or peer performance and the evaluation of a given performance against this reference. The choice of relevant reference can be guided by theory using axiomatic approaches, cf. e.g. Bogetoft and Hougaard(1999) and Färe and Lovell (1978), and it may be relevant to put some restrictions on the choice of reference. On the other hand, the choice of reference should also be a reflection of the preference of the user. It involves the trade-off between different performance dimensions and between different types of improvements. We shall therefore think of the reference selection as a choice problem. The user, whom we shall often think of as a decision maker (DM), chooses the benchmarking direction. To solve this choice problem, the user can be supported in many ways. He can basically draw on most of decision theory. In conformity with the flexibility philosophy and as a consequence of the many possible uses and users over time, we suggest to draw on the literature on *multiple criteria decision making* (MCDM). For text book introductions, see for example Bogetoft and Pruzan(1997) or Steuer(1986).

We realize of course that much of the benchmarking and efficiency analysis literature work on the no-preference-information assumption. In fact, this explains part of its success. On the other hand, focus on efficiency (doing things right) without being concerned about effectiveness (doing the right things) will in general lead to sub-optimal decisions.

The appropriate choice of a reference may not only be question of which direction to move in. It may also be a question of how far to move. In the short run, it may unrealistic to fully mimic a best practice unit. It is desirable therefore to allow some flexibility as to the choice of *performance level* against which to benchmark. The user shall for example be allowed to decide which best practice fractile to compare to.

In this paper we discuss in more details the design of such interactive, internet based benchmarking environments where the user specifies the desired mix of improvements as well as the performance level. We shall work both with parametric (statistical) as well as non-parametric (DEA) models. The user's specification of (relative) preference information affects the benchmarks and improvement potentials presented to him. We illustrate an implementation of a simple parametric model in 50 different industries at a commercial site. The model have capital and labor as inputs and gross profit as output. Furthermore a non-parametric DEA model is presented. 189 Danish banks are compared in an Internet based benchmarking model with 2 inputs and 3 outputs.

There are several papers that link multiple criteria decision making and benchmarking. By introducing information about desired improvements and trade-offs between alternative improvements, it is possible to beyond a simple efficiency analysis and towards a goal attainment or effectiveness analysis. Golany(1988a) and Ali, Cook and Seiford(1991) suggest the introduction of at least partial preference information in a dual formulation of the usual DEA models, while Golany(1988b) outlines the linkage with interactive multiple criteria methods. Along similar lines, Belton and Vickers(1992, 1993) suggest to integrate MCDM and DEA via a so-called VIDEA software, where the user can change the weights of the individual inputs and outputs. In the traditional DEA these weights are chosen such that the DMU being evaluated performs as good as possible. The most elaborate suggestions along these lines have come from Professor Korhonen and his coauthors. Korhonen and Laakso(1986) early introduced the so-called Reference Direction Approach to MCDM in a dynamic version supported by computer graphics. It was subsequently developed into the so–called *Pareto Race*, cf. Korhonen and Wallenius(1988). Pareto Race is an interface that support the user's search

4

on the frontier in a Multiple Objective Linear Programming (MOLP). Korhonen(1997) suggests the use of Pareto Race in DEA to choose a desired unit. Also, Joro et.al.(1998) emphasize the technical analogies between DEA and MCDM and Halme et.al. (1999) and Korhonen(2002) discuss the refinement of efficiency measures by incorporating preferences.

We deviate from the previous approaches linking multiple criteria decision making and benchmarking by stressing the individual learning and decision making perspective. For an unit seeking to improve, the interesting aspects to know are the improvement potentials and the possible trade-offs between alternative improvement dimensions. In the terminology of the literature, we are interested in the choice of a reference or peer unit and the comparison in absolute terms with this. We are much less interested in the measurement or index problem of summarizing the differences between the actual unit and the reference unit in a single number. For this reason, we draw on the MCDM literature because it contains useful knowledge about how to learn about and search among multiple dimensional alternatives.

The outline of the paper is as follows. In Section 2 we set the stage for benchmarking in general. Section 3 describes the interactive approach we propose. Integration with modern benchmarking techniques are discussed in Section 4 covering parametric and non-parametric benchmarking. In Section 5 , we describes the implementation of these ideas in Internet-based parametric and a non-parametric benchmarking modules. Section 6 discuss the two implementations and Section 7 concludes.

2   General benchmarking

Consider k organizations, usually referred to as Decision Making Unit (DMUs), that each transforms n inputs into m outputs. Let $x^i = (x_1^i, \ldots, x_n^i) \in \mathbb{R}_0^n$ be the inputs consumed and let $y^i = (y_1^i, \ldots, y_m^i) \in \mathbb{R}_0^m$ be the outputs produced by DMU^i, $i \in I = \{1, 2, \ldots, k\}$. The production possibility set is given by:

$$T = \{(x, y) \in \mathbb{R}_0^{n+m} | x \text{ can produce } y\} \tag{1}$$

The production correspondence is given by: $x \to P(x)$ and the consumption correspondence by: $y \to L(y)$, where

$$P(x) = \{y | (x, y) \in T\} \qquad L(y) = \{x | (x, y) \in T\} \tag{2}$$

i.e. $P(x)$ is the set of outputs that x can produce and $L(y)$ is the set of input that can produce y.

Inefficiency is the ability to reduce inputs without affecting output or the ability to expand output without requiring more inputs. In the multiple inputs, multiple outputs case a popular measure has become the so-called Farrell index. It measures the possibility to make proportional input reductions E or output expansions F:

$$E^i = \min\{E \in \mathbb{R}_0 | (Ex^i, y^i) \in T\} \tag{3}$$

$$F^i = \max\{F \in \mathbb{R}_0 | (x^i, Fy^i) \in T\} \tag{4}$$

Benchmarking compares the performance of a DMU to the frontier of T. Sometimes, it is useful to compare only towards a subset of T. The subset might exclude a certain percentage of the best performing DMUs, DMUs from certain geographic areas or DMUs of a certain sizes etc.

Figure 2 illustrate different benchmarking situations. In the ideal full information model at *Stage 1* the individual DMUs' effectiveness are measured by assuming that we know the preferences, $U(.)$, and the production technology, T. This full information approach is usually not feasible. Estimating utility functions are difficult and sometimes even theoretical impossible. Substituting the preferences with the criteria: "producing more with less" leads to the absolute efficiency model in *Stage 2*. Without a set of preferences a priori, we move from effectiveness to efficiency. That is, instead of a unique best plan, the result is a set of best performances - the efficient frontier. *Stage 2* still assume that we know the true technology, T. In most situations this is not the case. We therefore replace T with an estimate, T^*. At *Stage 3*, efficiency is measured relative to T^*, i.e. relative to the other DMUs. T^* can be determined by either parametric or non-parametric methods.

We reintroduce preferences in *Stage 4*. The true preferences $U(.)$ are approximated by a simplified preference model $U^*(.)$. Hereby relative efficiency is replaced with an approximation of relative effectiveness. The approximated preferences can be introduced via a communication process with an interactive exchange of preferences and benchmarks reflecting the submitted preferences.

We will distinguish between *individual benchmarking* and *overall benchmarking*. In individual benchmarking, the focus is on a detailed analysis of a single DMU, its improvement possibilities and the peer units corresponding to different improvement strategies. In overall benchmarking, the whole population of units are analyzed in terms of a common improvement perspective
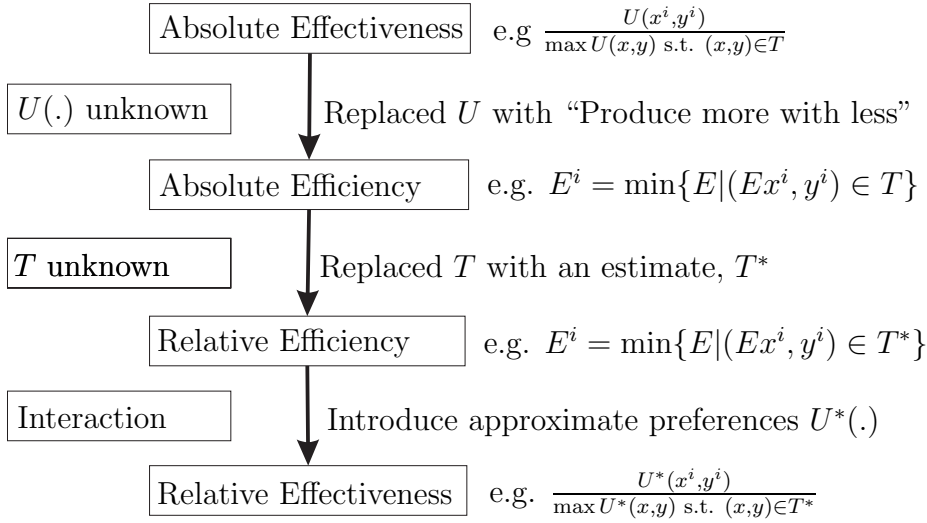
Figure 1: The basic benchmark approach

like the Farrell measure. Modern benchmarking studies using techniques like DEA have almost exclusively been used to evaluate the performance of all the units in an industry. We introduce the term individual benchmarking to emphasize also the usefulness of these techniques for detailed analyses of improvement possibilities in different directions for a single DMU. In the terminology of MCDM, both approaches are directed by the user - the Decision Maker (DM). Individual benchmarking is an analog to *progressive articulation of alternatives* approach, where the DM iteratively changes his preferences and hereby moves between benchmark. Overall benchmarking is a variety of the method of *prior articulation of alternatives*, where all alternatives are given to the DM.

3   Interactive benchmarking

We shall now leave the no-preference-information regime of traditional benchmarking and allow the user or decision maker DM to influence the benchmarking or reference selection.

In the context of individual benchmarking the starting point is typically the actual performance, $(x, y) = (x^i, y^i) \in T$, of a particular DMU. The user expresses his preferences by specifying directly or indirectly an appropriate benchmark $(\bar{x}, \bar{y})$ against which to compare (x, y). The issue now is what
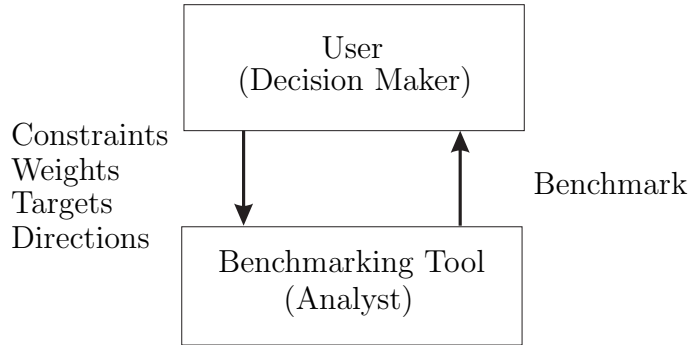
Figure 2: Progressive articulation of benchmarks

restrictions we can put on $(\bar{x}, \bar{y})$ and how we can support the user's choice of $(\bar{x}, \bar{y})$.

It is often natural to restrict the reference plan to be efficient. By definition $(\bar{x}, \bar{y})$ is *efficient* if there does not exist any point $(x^*, y^*) \in T$ such that:

$$x^* \leq \bar{x}, y^* \geq \bar{y} \qquad \text{and} \qquad x_i^* < \bar{x}_i \quad \vee \quad y_j^* > \bar{y}_j \qquad (5)$$

for some $i = 1, 2, \ldots, n$ or $j = 1, 2, \ldots, m$. That is, in an efficient plan, it is impossible to increase any output without decreasing other outputs or increase some inputs. Also, it is impossible to decrease any input without increasing other inputs or decreasing some outputs. Let T_E be the set of efficient plans in T.

The choice of a benchmark is essentially a multiple criteria decision problem. It involves a trade-off between different performance criteria, viz improvements in the different input and output dimensions. Many MCDM procedures can support the user's choice of benchmark. For a taxonomy of methods, see Bogetoft and Pruzan(1997). For a flexible benchmarking environment, we believe that it is most appropriate to use methods from the *progressive articulation of alternatives* class. In this class of MCDM methods, the user gradually learns about different alternative, here benchmarks. Also, he directs the search for new alternatives, i.e. benchmarks, via instructions to an analyst or computer code. The approach is illustrated in Figure 3.

A progressive articulation approach is attractive because it allows the user great flexibility in his learning. He is allowed to change his "preferences" as he go along and his implicit articulation of preferences is facilitated by the gradual revelation of actual benchmarks.

There are many methods based on the progressive articulation of alternative approach that can be relevant to apply. The user can direct the analyst or program using varying side constraints, weights as well as targets. We have experimented with all these methods in different applications. To illustrate the idea, however, it suffices to consider just one approach which we call *the directional approach*. This approach has proved useful in applications and it moreover links nicely with the modern benchmarking literature via the notion of a so-called directional distance functions, cf. Luenberger(1992) and Chambers, Chung & Färe(1995, 98).

In the directional approach, the user expresses his preferences or gives his instructions by specifying the direction, $d = (d_x, d_y) \in \mathbb{R}^{n+m}$, to look. DMU i's benchmark $(\bar{x}^i, \bar{y}^i)$ is hereafter given by $(x^i, y^i) + d \cdot \sigma$, where σ is:

$$\sigma = \max\{\sigma | (x^i, y^i) + d \cdot \sigma \in T\} \qquad (6)$$

By varying d, it is clearly possible to make any efficient production plan in T the desired benchmark for a given inefficient[2] DMU. Moreover the choice of d can be supported by thinking in terms of:

Side constraints: The user restricts certain inputs or outputs and a new point on the frontier, $(\hat{x}, \hat{y})$ that reflect these constraints are calculated[3]. In this case we can set $d = (\hat{x}, \hat{y}) - (x^i, y^i)$.

Weights: The user submits relative weights between inputs and outputs, w. These weights can be thought of as a subjective price vector. In this case we can use $d = w$.

Targets: The user submits a goal, (x^g, y^g), and we use $d = (x^g, y^g) - (x^i, y^i)$.

It might be useful to bound the directional vector to ensure that the resulting reference plan is efficient. In some applications it might also be relevant to restrict the search for benchmarks to points that are *weakly improving* in all dimensions i.e. to require

$$0 \geq d_x \qquad \text{and} \qquad 0 \leq d_y \qquad (7)$$

---

[2]Starting out at an efficient point, the procedure can not generate all other efficient points. In this case it will be sufficient to let the search start out at a slightly pertubed point. We will return to this in Section 4 below.

[3]There are many ways to pick $(\hat{x}, \hat{y})$. We can for example maximize $\sum_{i=1}^{m}(\hat{y}_i - y_i) + \sum_{i=1}^{n}(x_i - \hat{x}_i)$ subject to the conditions that $(\hat{x}, \hat{y}) \in T$ and that $(\hat{x}, \hat{y})$ fulfill the side constraints.

where d_x and d_y are the n and m dimensions of d corresponding to the inputs and outputs.

Unlike individual benchmarking the purpose of overall benchmarking is not to search the frontier, but to present a ranking of all DMUs. Traditionally all DMUs are ranked using the Farrell measure. One way to incorporate preferences would be to measure efficiency for all DMUs using the same direction d rather than the units specific directions implied by the Farrell approach. By changing d, the ranking of all DMUs may change. We shall return to the choice of ranking index in Section 4.

4 Parametric and non-parametric benchmarking

In the discussion so far, we have assumed that the underlying technology T was given. In practice, T must be estimated from observed performances, say the inputs $x^i = (x_1^i, \ldots, x_n^i) \in \mathbb{R}_0^n$ consumed and outputs $y^i = (y_1^i, \ldots, y_m^i) \in \mathbb{R}_0^m$ produced in $\text{DMU}^i, i \in I$. Much of the progress in modern benchmarking theory has been on the estimation of advanced production structures.

It is common to distinguish between parametric and non-parametric approaches.

In the *parametric approach*, initial regularity on T is by postulating a certain functional structure, say

$$T = \{(x, y) | f(x, y, \alpha) \leq 0\}$$

where f is a function mapping inputs x, outputs y and parameters α into the real numbers. By estimating α from observations of realized input-output combinations, an approximation of T is obtained that can be used as the basis for reference point selection. In the case of a single output (classical production function) or a single input (classical cost function), standard econometric theory can be used to estimate α. Also, more advanced Stochastic Frontier Analysis (SFA) can be used. These methods work with both a normally distributed noise term that can increase or decrease production (or costs) in the usual way, and a one-side inefficiency term, that can only decrease production or increase costs. For an introduction to SFA, see for example Coelli et.al. (1998).

Now, given a functional representation of T the directional approach requires the analyst or computer program to find DMU i's benchmark $(\bar{x}^i, \bar{y}^i) = (x^i, y^i) + \sigma^* \cdot d$, where σ^* is:

$$\sigma^* = \max\{\sigma | f(x^i + \sigma \cdot d_x, y^i + \sigma \cdot d_y, \alpha) \leq 0\} \tag{8}$$

In general, numerical methods will be needed to solve this problem, but if the functional form is sufficiently well behaved or the number of input-output dimensions sufficiently small, analytical solutions may be possible as well.

In the non-parametric approach, the technology T is estimated from a set of basic postulates about T and the so-called minimal extrapolation principle. The basic postulates about T are usually:

Free disposability: $(x', y') \in T$ and $x'' \geq x'$ and $y'' \leq y' \Rightarrow (x'', y'') \in T$ i.e. "more can produce less".

Convexity: T is convex. That is weighted averages of feasible production plans are feasible as well. This is usually the case, although in the case of non-parametric models one might relax this assumption.

Return to scale: $(x', y') \in T \Rightarrow s(x', y') \in T$ for $s \in S(h)$, where h equals: crs (constant return to scale), drs (decreasing return to scale), vrs (variable return to scale) or irs (increasing return to scale), and $S(crs) = \mathbb{R}_0, S(drs) = [0, 1], S(vrs) = 1$ or $S(irs) = [1, \infty)$.

The idea of *minimal extrapolation* is now to find the smallest subset of $\mathbb{R}_0^{n+m}$ that contains the actual input-output observations and satisfy certain combinations of the assumptions above. The original DEA model by Charnes, Cooper and Rhodes(1978,79) invokes free disposability, convexity and constant return to scale. For alternative specifications invoking different combinations of assumptions, see for example Charnes, Cooper, Lewin and Seiford (1994). In general, the technology estimated by the non-parametric approach can be expressed using mathematical constraints. Thus, for example, invoking the free disposability and convexity assumption leads to the estimate

$$T = \{(x, y) | x \leq \sum_{j \in I} \lambda^j x^j, y \geq \sum_{j \in I} \lambda^j y^j, \sum_{j \in I} \lambda^j = 1, \lambda^j \geq 0 \forall j \in I\}$$

and if we assume decreasing return to scale or constant return to scale, we simply replace $\sum_{j \in I} \lambda^j = 1$ with either $\sum_{j \in I} \lambda^j \leq 1$ or $\sum_{j \in I} \lambda^j \in \mathbb{R}_0$.

Given a non-parametric representation of T as above, the directional approach requires the analyst or computer program to find DMU i's benchmark $(\bar{x}^i, \bar{y}^i) = (x^i, y^i) + \sigma^* \cdot d$, where σ^* is a solution to the following linear programming (LP) program

$$\sigma^* = \max_{\sigma, \lambda} \sigma$$
$$\text{s.t.} \quad x^i \geq \sum_{j \in I} \lambda^j x^j - \sigma d_x$$
$$y^i \leq \sum_{j \in I} \lambda^j y^j - \sigma d_y \tag{9}$$
$$\sum_{j \in I} \lambda^j = 1$$

Whether we use a parametric or a non-parametric approach, the traditional Farrell approach is a special case of the directional approach. With $d = (-x^i, 0)$ we have $E = 1 - \sigma$ in the input based case and with $d = (0, y^i)$ we have $F = 1 + \sigma$ in the output based case.

The estimated inefficiency for DMU^i, $\sigma^* d$, is an inefficiency measure in absolute numbers. This has advantages in the case of individual benchmarking. However moving from individual to overall benchmarking a relative performance index is needed to *rank all units*. In Figure 3 an index for a pessimistic improvement potential is given by:

$$\min \left(1 - \frac{\bar{x}_h}{x_h^i}, h = 1, \ldots, n, 1 - \frac{y_k^i}{\bar{y}_k}, k = 1, \ldots, m. \right) \tag{10}$$

The interpretation of the improvement potential is straightforward: it gives the minimum each and every inputs can be reduced and outputs expanded compared with the benchmark. An efficient firm will have 0 improvement potential. In the case of a Farrell measure the improvement potential would equal $1 - E$ (input) and $F - 1$ (output).

The improvement index is easy to interpret, however it is very sensitive to the DMU's actual structure and it is not an ideal ranking index. A more appropriate ranking index, which however is less intuitive, is given in Figure 4. This index is entirely built on distances. Although it does not have the same easy interpretation it is identical with the improvement index in the case of Farrell.
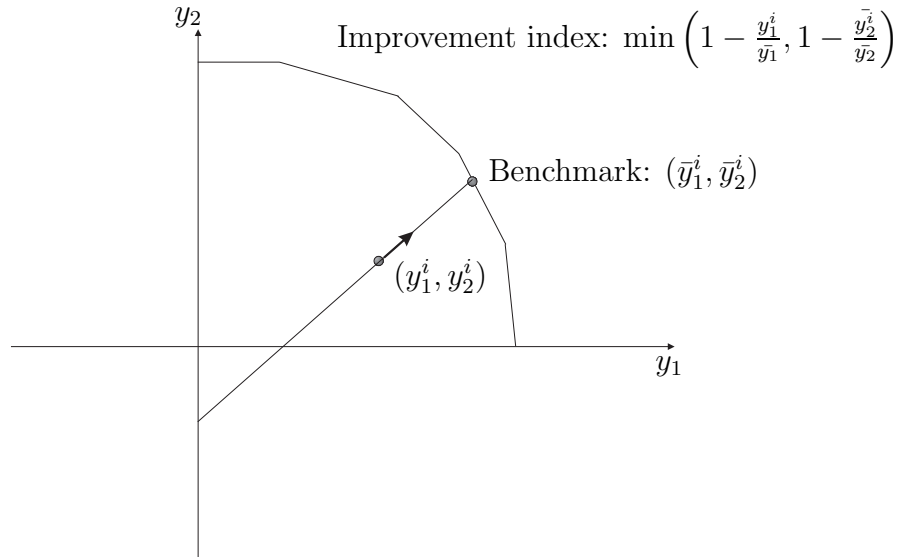
Figure 3: Index for improvement potential
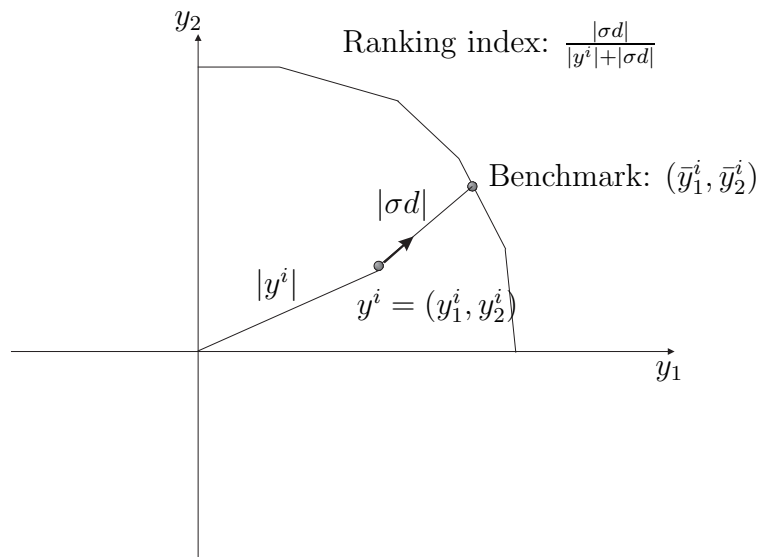
Figure 4: Ranking index

5 Applications

In this section, we describe the implementation of two interactive, internet-based benchmarking models. The purpose of the first implementation is to illustrate the strength of interactive benchmarking with a simple parametric model. The second implementation shows how a traditional non-parametric DEA model can be expanded to an Internet based benchmarking model.

5.1 Parametric application

The software described in this section is used at the commercial Internet site: www.managershotline.dk, that sells managerial advises. 50 different industries in Denmark are covered. In each industry a simple parametric *capital-labor-gross profit* model are estimated. The data are provided by Købmandsstandens Oplysningsforbund.

The variables are:

Labour Number of full-time employees (L)

Capital Fixed assets, the part of all assets that is continuously owned (C)

Gross profit Total turnover excluding taxes minus primary inputs (Y)

These numbers are easily found in most accounts. The model is a simple Cobb-Douglas function, that explain the gross profit by labor and capital:

$$Y = \beta_0 L^{\beta_1} C^{\beta_2} \tag{11}$$

After a logarithmic transformation the β parameters were estimated using Ordinary Least Square. The statistical tests were supporting the model e.g. the R^2 measures belong to the interval: $[0,62; 0,92]$[4].

---

[4]Except the industry "Dentists" with a R^2 equal to 0,48.

The benchmarking

The user submits data on labor and capital and receive 3 estimated levels of gross profit:

- The expected gross profit $\hat{Y}$

- The 25 % best gross profit Y_{25}

- The 10 % best gross profit Y_{10}

Here, $\hat{Y}(L^i, C^i)$ is the expected gross profit given the user's labor and capital numbers, i.e. $\hat{Y}(L^i, C^i) = \beta_0 L^{i\beta_1} C_2^{i\beta_2}$ with the estimated parameter values inserted.

The next two benchmarks are found by scaling the expected gross profit with the corresponding efficiency score fractiles. The inverse Farrell based efficiency score is

$$G^i = \frac{Y^i}{\hat{Y}(L^i, C^i)} \tag{12}$$

Let G_{25} be the 25 % highest efficiency score G^i. Then in the 25 % best gross profit scenario, the DMU^i is compared with:

$$Y_{25} = G_{25} \cdot \hat{Y}(L^i, C^i) \tag{13}$$

and similar for Y_{10}.

Surfing the frontiers

After the initial benchmark the user can study the firms improvement potential in more details. The kind of questions that can be answered are "How many employees and how much fixed capital could I save if I was as efficient as the 25 % best".

The same level of gross profit can be produced with different combinations of L and C. The possibility of substitution makes it useful to explore the frontiers. The traditional benchmark that comes from a proportional (Farrell) change in the variables is used as the starting point:

$$Y^i = G_{25}\beta_0(\delta \cdot L^i)^{\beta_1}(\delta \cdot C^i)^{\beta_2} \tag{14}$$

The proportional reduction δ and the saving potential in real numbers given the ambition to do as well at the chosen fractile, here the 25% best, are given as a start. From this point the user can freely move along the frontier. The user is simply submitting a new level of either L or C and the corresponding values of C or L are calculated.

Figure 5 shows a situation, where the user perform between the 25 % and the 10 % best. The figure depicts 3 iso-efficiency curves, all points produce the same output, Y^i, with different efficiency levels. If the user chose to benchmark against the 10 % best, point A would be the starting point. A represent a proportional reduction in all inputs. He can now move along the frontier by changing either L or C^5. The user could also compare himself with the 25 % best, which initially will bring him to point B. Point B represent a proportional increase in all inputs required for producing the same with this lower efficiency.
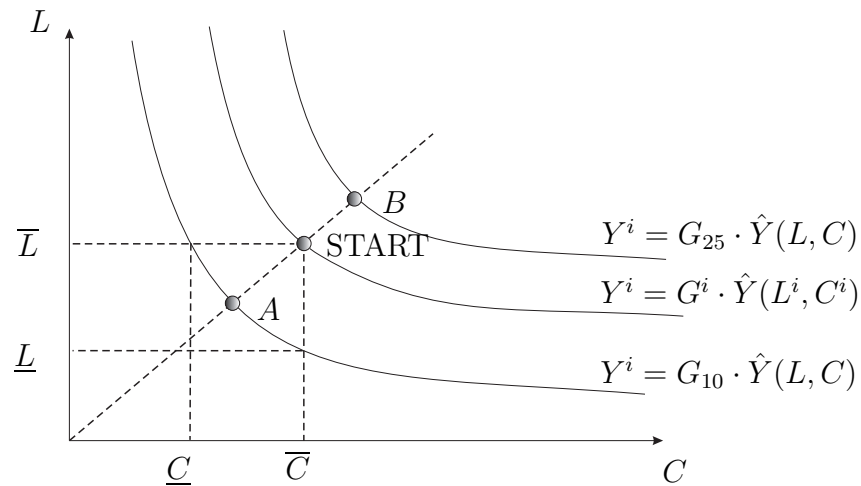
Figure 5: Choosing different frontiers

Figure 6 is a screen shot from the searching part. Figure 6 corresponds to point A in Figure 5 and it illustrate how the search can be communicated.

---

[5]It is basically up to the user to decide to what degree substitution between L and C is possible but to reflect the estimation conditions, we have introduced upper and lower limits $(\underline{L}, \overline{L})$ and $(\underline{C}, \overline{C})$ in the program.
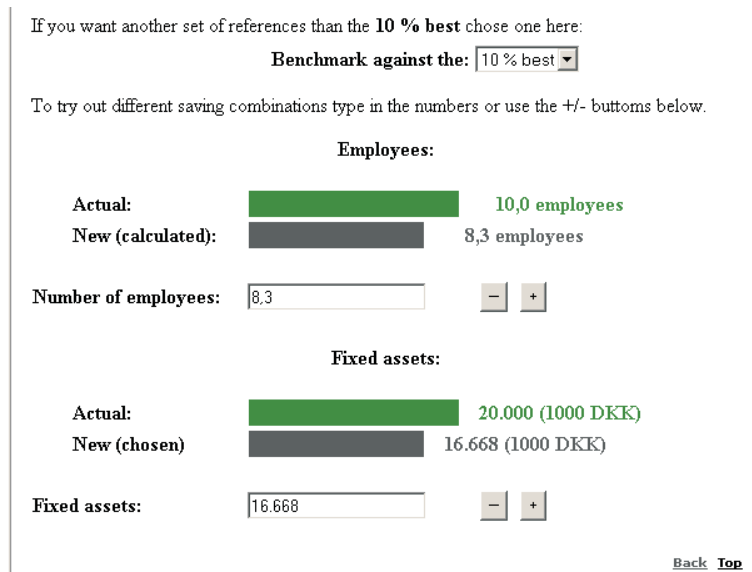
Figure 6: Searching the frontier (screen shot)

5.2 Non-parametric application

This Section describes the design of an Internet based interactive benchmarking system based on a non-parametric DEA model. The software have been programmed in SAS and SASIntr is used to create a CGI-connection (common gateway interface) through a browser. The setup is a two tier system where all calculation and data is managed through SAS.

We have analyzed 189 Danish commercial and savings banks in a DEA model with two inputs:

- **Staff & admin**, Staff and administrative expenses and other operating expenses.

- **Own funds**, Own funds total.

and three outputs:

- **Net income (Interest)**, Net income from interest.

- **Charges a.o. income**, Charges and commissions receivable (- payable) and other operating income.

17

- **Guaranties etc.**, Guaranties and other commitments.

The production possibility set, T, is assumed to be convex and free disposable. This leads to a cautious, small envelopment of the data. Benchmarks are calculated using the directional distance approach.

The user can either choose individual benchmarking or overall benchmarking. In individual benchmarking the user searches the frontier interactively by changing the direction d. In overall benchmarking, all DMUs are ranked relatively to each other and the result is presented by sorted lists and plots. The user can move from individual benchmarking to overall benchmarking at any time.

Individual benchmarking

Figure 7 shows the user interface in individual benchmarking. The benchmark given in Figure 7 is determine by the submitted "weights". The weights point out a direction d in which the best $(x, y) \in T^*$ is given. The side constraints can be used to further restrain the search. They are initially given with default values equal to the extreme observations in the data set. The improvement potential 1,8 % is the smallest feasible proportional contraction of all inputs and expansion of all outputs with this benchmark.

The benchmark is a linear combination of a set of efficient peer DMUs. In this case the benchmark consist of 4 different banks. The peers actual data is given in more detail by clicking the "Peers in details" button. At the bottom of the page the actual performance and the reference peer performance are illustrated graphically.

The search starts at the DMU^i's actual performance (x^i, y^i). Using positive weights corresponding to $d_x < 0$ and $d_y > 0$, the system will give benchmarks that weakly dominate the actual performance. However, by allowing negative weights an (inefficient) unit can search the entire frontier by changing the weights. As fixed points the proportional Farrell projections are made available via single clicks.

To measure the improvement potential, it is natural to start out at the actual performance. However, this may prevent an efficient DMU from exploring the entire frontier. The problem is that points at the relative interior of the facet to which the unit belong will not be generated using the directional search program from the previous section. To allow efficient units to search
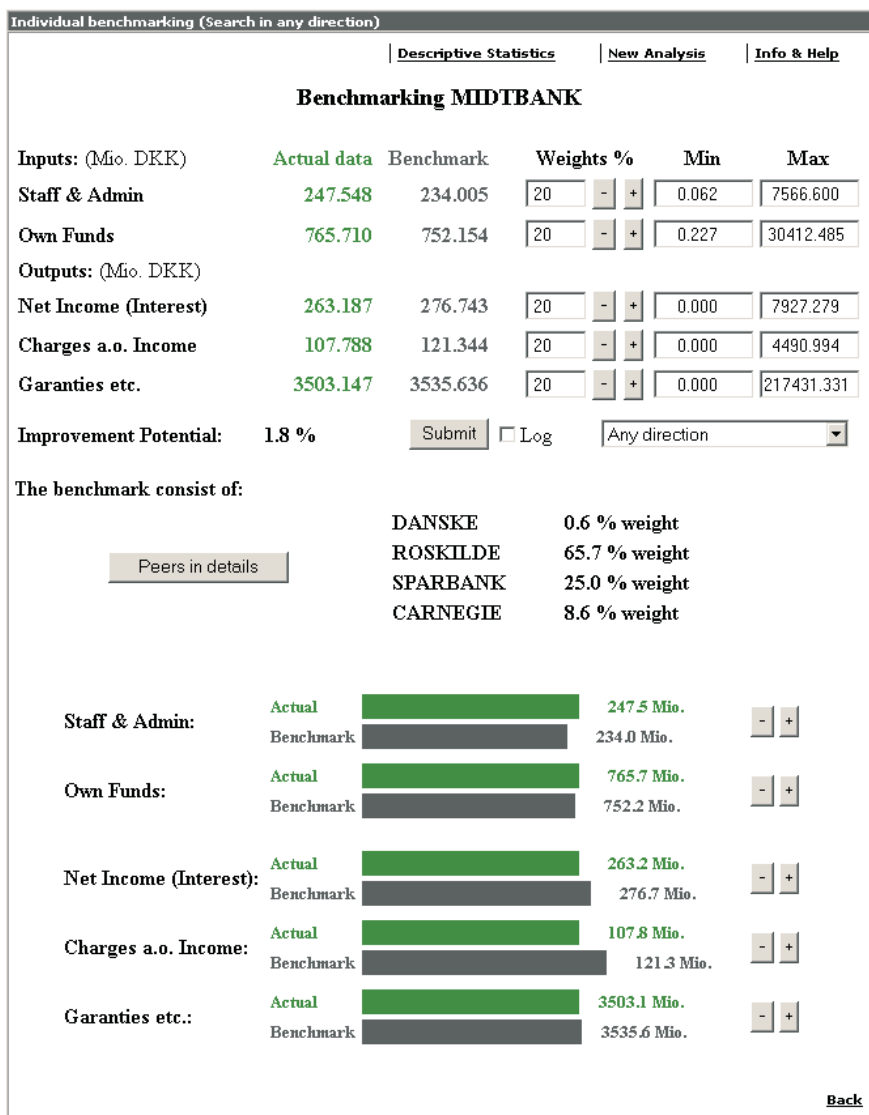
| | Descriptive Statistics | New Analysis | Info & Help |

Benchmarking MIDTBANK

Inputs: (Mio. DKK)	Actual data	Benchmark	Weights %			Min	Max
Staff & Admin	247.548	234.005	20	-	+	0.062	7566.600
Own Funds	765.710	752.154	20	-	+	0.227	30412.485
Outputs: (Mio. DKK)							
Net Income (Interest)	263.187	276.743	20	-	+	0.000	7927.279
Charges a.o. Income	107.788	121.344	20	-	+	0.000	4490.994
Garanties etc.	3503.147	3535.636	20	-	+	0.000	217431.331

Improvement Potential: 1.8 %   Submit   ☐ Log   Any direction ▼

The benchmark consist of:

Peers in details

DANSKE	0.6 % weight
ROSKILDE	65.7 % weight
SPARBANK	25.0 % weight
CARNEGIE	8.6 % weight

	Actual / Benchmark		
Staff & Admin:	Actual	247.5 Mio.	- +
	Benchmark	234.0 Mio.	
Own Funds:	Actual	765.7 Mio.	- +
	Benchmark	752.2 Mio.	
Net Income (Interest):	Actual	263.2 Mio.	- +
	Benchmark	276.7 Mio.	
Charges a.o. Income:	Actual	107.8 Mio.	- +
	Benchmark	121.3 Mio.	
Garanties etc.:	Actual	3503.1 Mio.	- +
	Benchmark	3535.6 Mio.	

Back

Figure 7: Interface for Individual Benchmarking

19

the frontier - and to give almost efficient units a more smooth search, the "free search" module can be used. If the user select "free search" the frontier can be searched starting at a strongly inefficient point. The starting point uses the double of all inputs to produce half of all outputs as actual DMU does.

Overall benchmarking

Overall benchmarking list the relative performance of all DMUs sorted by the score. The score could be the well-known Farrell input or output oriented scores or it could be the directional score presented in Section 4. The distribution of the scores are provided in different diagrams like in a traditional DEA study. Furthermore the peer-structure of all DMUs are provided on request.

At any time, the user can chose any DMU and go to individual benchmarking to explore the improvement potential of this DMU. Furthermore the user can drill down to the very details of the individual DMUs or get descriptive statistics for the entire data set.

6   Discussion

Although both the parametric and the non-parametric approaches can handle multiple inputs and multiple outputs models, they have different cost and benefit profiles.

Outliers are not too troublesome to a - well performed - parametric estimation. There are ways also to deal with them in DEA, e.g. peeling, sensitivity analysis, more frontiers etc. Still, the quality of data is more important in DEA.

Little a priori information about the underlying technology is needed in a non-parametric model compared to a parametric model. In terms of information to the user the non-parametric approach also provides more detailed data driven information. Information on the performance of actual peer units is very useful and in high demand. Anonymity requirements, however, may work in the opposite direction. A parametric representation provides the highest level of anonymity. A fair level of anonymity could be reach in non-parametric models as well, e.g. new efficient DMU can be created using sampling techniques to cover the actual DMUs.

A parametric representation is more complicated to establish than a non-parametric one, but ones established it is easier to use. In terms of computation time parametric models are more sensitive to the number of variables and less depended on the number of DMUs than a non-parametric approach.

7 Conclusion

In this paper, we have proposed to embed modern benchmarking techniques in interactive benchmarking environments that can be assessed via the internet. We believe that this will increase the usefulness of these techniques since it will allow individualized analyzes and support learning that is directed by the learning unit and not by an analyst. Ultimately, such systems most prove there worth in real life applications. From other computer implementations of decision support systems etc, however, it is well known that the communication between the program and the user, including the use of an appealing design of the interface, is crucial. We have - starting from the theory and practice of multiple criteria decision making - made several suggestions about the design of such easy to use and simple interfaces.

Several issues remains to be solved. We have already discussed outliers, information and anonymity and computation time. A fourth issue concerns the use of the direction approach in overall benchmarking. We need a better ranking index with an intuitive interpretation. Also, the computation time must be reduced when applying a certain direction to all units at the same time.

References

ALI, A., W. COOK, AND L. SEIFORD (1991): "Strict vs. Weak Ordinal Relations for Multipliers in Data Envelopment Analysis," *Management Science*, 37, 733–738.

BELTON, V., AND S. P. VICKERS (1993): "Demystifying DEA - A Visual Interactive Approach Based on Multiple Criteria Analysis," *Operation Research Society*, 44(9), 883–896.

BOGETOFT, P. (1994): "Incentive Efficient Production Frontiers: An Agency Perspective in DEA," *Management Science*, 40, 959–968.

———— (2000): "DEA and Activity Planning under Asymmetric Infromation," *Management Science*, 13, 7–48.

BOGETOFT, P., AND J. L. HOUGAARD (1999): "Efficiency Evaluation Based on Potential (Non-proportional) Improvements," *Journal of Productivity Analysis*, 12, 231–245.

BOGETOFT, P., AND P. PRUZAN (1997): *Planning with Multiple Criteria.* Copenhagen Business School Press, 2 edn.

CHAMBERS, R. G., Y. CHUNG, AND R. FÄRE (1995): "Benefit and distance functions," *Journal of Economic Theory.*

———— (1998): "Profit, Directional Distance Functions, and Nerlovian Efficiency," *Journal of Optimization Theory and Application*, 2, 351–364.

CHARNES, A., W. W. COOPER, A. LEWIN, AND L. M. SEIFORD (1994): *Data Envelopment Analysis: Theory, Methodology and Application.* Kluwer Academic Publishers.

CHARNES, A., W. W. COOPER, AND E. RHODES (1978): "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, 2, 429–444.

———— (1979): "Short Communication: Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, 3, 339.

COELLI, T., D. P. RAO, AND G. E. BATTESE (1998): *An Introduction to Efficiency and Productivity Analysis.* Kluwer Academic Publisher.

EMROUZNEJAD, A. (2001): "An Extensive Bibliography of Data Envelopment Analysis Volume I to V," http://www.warwick.ac.uk/ bsrlu/.

FÄRE, R., AND C. LOVELL (1978): "Measuring the technical efficiency of production," *Journal of Economic Theory*, 19, 150–162.

GOLANY, B. (1988a): "An Interactive MOLP Procedure for Extension of DEA to Effectiveness Analysis," *Journal of Operational Research Society*, 39, 725–734.

———— (1988b): "A Note on Including Ordinal Relations Among Multipliers in Data Envelopment Analysis," *Management Science*, 34, 1029–1033.

HALME, M., T. JORO, P. KORHONEN, S. SALO, AND J. WALLENIUS (1999): "A value efficiency approach to incorporating preference information in data envelopment analysis," *Management Science*, 45, 103–115.

JORO, T., P. KORHONEN, AND J. WALLENIUS (1998): "Structural comparison of data envelopment analysis and multiple objecive linear programming," *Management Science*, 44, 962–970.

KORHONEN, P. (1997): "Searching the Efficient Frontier in Data Envelopmant Analysis," Interim Report IR-97-79, International Institute for Applied Systems Analysis.

KORHONEN, P., H. MOSKOWITZ, AND J. WALLENIUS (1992): "Multiple criteria desicion support - A review," *European Journal of Operational Research*, 63, 361–375.

KORHONEN, P., M. SOISMAA, AND A. SILJAMAKI (2002): "On the Use of Value Efficiency Analysis and Some Further Developments," *Journal of Producticity Analysis*, 17, 49–65.

KORHONEN, P., AND J. WALLENIUS (1988): "A Pareto Race," *Naval Research Logistics*, 35, 615–623.

LUENBERGER, D. (1992): "Benefit functions and duality," *Journal of Mathematical Economics*, 21, 461–481.

PEKKA KORHONEN, M. S. (2001): "Ressource Allocation Based on Efficiency Analysis," Working Paper, Helsinki School of Economics.

STEUER, R. (1986): *Multiple Criteria Optimization: Theory, Computation and Application.* Wiley.