

I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN



D I S C U S S I O N
P A P E R

2014/16

Efficiency and benchmarking with
directional distances: a data driven approach

DARAIO C. AND L. SIMAR

EFFICIENCY AND BENCHMARKING WITH DIRECTIONAL DISTANCES: A DATA DRIVEN APPROACH

CINZIA DARAIIO* LÉOPOLD SIMAR*,§
daraio@dis.uniroma1.it leopold.simar@uclouvain.be

April 23, 2014

Abstract

In efficiency analysis the assessment of the performance of Decision Making Units (DMUs) relies on the selection of the direction along which the distance from the efficient frontier is measured. Directional Distance Functions (DDFs) represent a flexible way to gauge the inefficiency of DMUs. Permitting the selection of a direction towards the efficient frontier is often useful in empirical applications. As a matter of fact, many papers in the literature have proposed specific DDFs suitable for different contexts of application. Nevertheless, the selection of a direction implies the choice of an efficiency target which is imposed to all the analyzed DMUs. Moreover, there exist many situations in which there is no a priori economic or managerial *rationale* to impose a subjective efficiency target. In this paper we propose a data-driven approach to find out an “objective” direction along which to gauge the inefficiency of each DMU. Our approach permits to take into account for the heterogeneity of DMUs and their diverse contexts that may influence their input and/or output mixes. Our method is also a data driven technique for benchmarking each DMU. We describe how to implement our framework and illustrate its usefulness with simulated and real datasets.

Keywords: DEA, benchmarking, directional distance functions, nonparametric estimation, heterogeneity, performance, productivity, organizational studies.

JEL Classification: C14, C40, C60, D20

*Department of Computer, Control and Management Engineering Antonio Ruberti (DIAG), University of Rome “La Sapienza”, Rome, Italy.

§Institute of Statistics, Biostatistics et Actuarial Sciences, Université Catholique de Louvain, Louvain-la-Neuve, Belgium. Research supported by IAP Research Network P7/06 of the Belgian State (Belgian Science Policy).

1 The Set-up

Data Envelopment Analysis (DEA) and related approaches like Free Disposal Hull (FDH) are very popular techniques used in Operations Research (OR) and Management Science (MS) to evaluate the performance of Decision Making Units (DMUs) operating in a given sector of activity. The economic underpinnings of these approaches dates back to the work of Koopmans (1951) and Debreu (1951) on the activity and productivity of firms. In a given sector of activity, DMUs combine/transform a set of inputs to produce a set of outputs. In an efficiency analysis, usually the objective is first to consider the space of technically attainable combinations of inputs and outputs and then to measure the inefficiency of the units by their distance to the optimal frontier of the attainable set. A nice and detailed presentation of the economic theory of production can be found, e.g. in Shephard (1970).

Farrell (1957) is the first empirical analysis of productive efficiency. It defines as an estimate of the attainable set, the free disposal convex closure of the cloud of observed inputs and outputs, in a sample of DMUs. The OR and MS literature popularized the techniques, and several variants, with linear programming techniques, as in Charnes, Cooper and Rhodes (1978), under the heading of DEA. The procedure was extended to non convex attainable sets by Deprins et al. (1989) introducing the FDH estimators. See e.g. Emrouznejad et al. (2008) for a survey of the many applications of DEA in OR/MS.

Most of the original approaches were in the spirit of Debreu and Farrell and the distance of an observed DMU to the efficient frontier was assessed in a “radial” way either in the input space or in the output space. For instance the input oriented inefficiency score of a given DMU was measured by the maximal proportionate reduction of the inputs that is feasible to reach the efficient boundary, given its level of production of the outputs.

After the introduction of Directional Distance Functions (DDF) in the efficiency analysis context (Chung, Färe and Grosskopf, 1997 and Chambers, Chung and Färe, 1998; Färe and Grosskopf, 2000), the literature has proposed several directions along which to measure the inefficiency and investigated their properties. Different fields of applications have used these flexible DDF e.g. in banking, education, mutual funds (e.g., Kerstens, Mounir and Van de Woestyne, 2012) and in general also, regulated sectors in which DMUs could complain about differences in environmental/contextual conditions. The great flexibility is particularly appreciated when some inputs or outputs are non-discretionary, or not under the control of the manager. This is also the case when some inputs or outputs are exogenously fixed (see Banker and Morey, 1986, for practical examples). All these cases are handled by choosing a direction in the input-output space that does not involve these particular components (see Daraio and Simar, 2014 for details on how to handle these cases in practice).

The definition of a direction along which to measure the distance of DMUs from the efficient frontier is really crucial as it affects the efficiency scores and the evaluation of the performance (see e.g. Peyrache and Daraio, 2012 or Kerstens et al., 2012). The choice of a direction implies the selection of an efficiency target. The price to pay is that the efficiency investigation loses its objectivity and becomes subjective, according to the direction selected by the analyst. In some cases, the regulator (see e.g. Bogetoft, 1997 on the use of efficiency analysis tools in a regulatory framework) or the policy maker may have an efficient target or a set of preferences translated in indicators that wants to impose to the analyzed firms (see e.g. Korhonen, Tainio and Wallenius, 2001). However, in a lot of situations policy makers and managers may be interested in avoiding the imposition of targets or paths to the DMUs they are monitoring and hence they may be willing to search for an “objective direction” along which to compare their performance.

Another problem in this context could be that the analysed units could refuse to be compared with very different units and may ask for a comparison that takes into account their specific characteristics. There may therefore be a need to avoid comparing units that are very different from each other (compare apples with oranges) and then to take into account the heterogeneity of the units in the comparison of their performance.

Long debates have been discussed in the economic literature on which direction should be chosen, and many choices are possible, see e.g. Färe, Grosskopf and Margaritis (2008). On the one hand, we can choose an individual direction specific to each DMU, like e.g. the direction given by the input (or the output) mix. In such cases, if we keep the outputs fixed (resp. or the input fixed) we can recover the Farrell-Debreu radial inefficiency scores. On the other hand, at the other extreme we can choose a common direction for all the DMUs, called the “egalitarian” strategy in Färe et al. (2008). An example of such common direction is to take the average input mix and the average output mix, if no particular efficiency target is selected. We come back to the latter below for suggesting alternate approaches.

When market prices are observed and firms have a profit maximizing behavior, Zofio, Pastor and Aparicio (2013) suggest to choose the directional distance which projects inefficient DMUs towards profit maximizing benchmarks and Färe, Grosskopf and Whittaker (2013) propose an endogenous choice of the direction obtained by searching the (normalized) direction which maximizes the output (in an output-oriented framework).

In this paper we propose an automatic data-driven approach which permits the selection of *context specific (or local)* directions of DMUs with similar exogenous conditions without any assumption on the behavior of DMUs (no profit maximizing or cost minimizing behavior, no information on prices needed). The proposed approach offers a threefold contribution.

Firstly it offers the opportunity to select a “context-dependent (or local)” direction to-

wards the efficient frontier according to the closeness (or based on the proximity) of the production mix or of the contextual conditions (factors) faced by the DMUs. The localization can be done according to exogenous or external (contextual) factors (*localization on external conditions*) or on the inputs if non discretionary inputs are considered (*localization on inputs rigidity*) or on the outputs if non-discretionary outputs are considered (*localization on outputs rigidity*). We introduce heterogeneity factors that may play a role on the rays or production mix. We will call these heterogeneity factors as $W \in \mathbb{R}^d$ hereafter. The procedure will be able to detect irrelevant factors.

Secondly, it allows a reasonable automatic selection of the direction among the two extreme cases: a unique (global average) direction for all DMUs (the empirical *egalitarian* approach), or a different direction for each DMU (*individual-specific* approach). By doing this we reformulate what would be a reasonable empirical egalitarian direction, addressing some of the issues linked to the traditional definition of the average directions.

Thirdly, the procedure produces as a byproduct an automatic “peer grouping” of DMUs with comparable external-environmental conditions and similar levels of their production mix without having to carry out a clustering exercise in which the choice of the number of clusters is arbitrary. This allows us to carry out a data-driven benchmarking analysis of DMUs, paying attention to comparison with comparable units.

From a technical point of view, we will employ up-to-date nonparametric regression on angular data, analyzed recently in Di Marzio, Panzera and Taylor (2013) to determine the local averages of the directional vectors.

The paper is organized as follows. In Section 2 we summarize the main concepts for directional distances and their estimators and we present the basics for nonparametric regression on angular data in a simplified way (the technical details are given in Appendix B). Section 3 illustrates the advantages of our approach from a benchmarking point of view. Section 4 describes all the practical methodological steps of the procedure and Section 5 reports some illustrations on simulated and real datasets. Section 6 concludes the paper. Appendix A clarifies the notations we use for multidimensional polar coordinates.

2 Directional Distances Functions and Searching for Directions

2.1 Directional DEA

We concentrate our presentation on DEA but all the methodology is valid and can be applied to FDH, robust and conditional efficiency measures (for an overview, see Daraio and Simar,

2007). We consider the case where DMUs produce a vector of outputs $y \in \mathbb{R}^q$ by combining inputs $x \in \mathbb{R}^p$.¹ We will denote by Ψ the attainable set, which is the set of (x, y) technically achievable:

$$\Psi = \{(x, y) \in \mathbb{R}^{p+q} | x \text{ can produce } y\}. \quad (2.1)$$

Given a directional vector $d_x \geq 0$ for the inputs and a directional vector for the outputs $d_y \geq 0$ the DDF of a unit operating at level (x, y) is defined as (see Chambers et al. 1996, 1998)

$$\delta(x, y; d_x, d_y) = \sup\{\beta > 0 | (x - \beta d_x, y + \beta d_y) \in \Psi\}, \quad (2.2)$$

indicating that we measure the distance of unit (x, y) from the efficient boundary of Ψ in an additive way along the path defined by $(-d_x, d_y)$. Obviously when $d_y = 0$ and $d_x = x$ if $x > 0$, we recover the input-oriented efficiency measure of (x, y) (the same for the output orientation when $d_x = 0$ and $d_y = y$ if $y > 0$).

Note that, as pointed out and proven in the literature (see e.g. Appendix A in Simar and Vanhems, 2012), when the units of the directional vectors are the same as the units of the inputs/outputs, the DDF is unit free in the following sense, for all $a \in \mathbb{R}_+^p$ and all $b \in \mathbb{R}_+^q$ we have $\delta(x, y; d_x, d_y) = \delta(a.*x, b.*y; a.*d_x, b.*d_y)$, where $.*$ denotes the Hadamard product between vectors (component wise). This is particularly useful when the units of the components of x and/or of y are quite different.

Sometimes researchers prefer to work with normalized distances, i.e. such that $\|d\| = 1$ where $d' = (d'_x, d'_y) \in \mathbb{R}^{p+q}$ (see e.g. Färe et al., 2008). This has the effect of rescaling the DDF by the length of d . To be explicit, denoting $\tilde{d}_x = d_x/\|d\|$ and $\tilde{d}_y = d_y/\|d\|$ we have $\delta(x, y; \tilde{d}_x, \tilde{d}_y) = \|d\|\delta(x, y; d_x, d_y)$. The advantage is that this measure gives directly the euclidean distance between (x, y) and its target on the efficient frontier, but the drawback is that this measure of DDF is no more unit free. Both approaches can be used, but the interpretation of the DDF has to be adapted.

These distances can be estimated from a sample of observed units $\mathcal{X} = \{(X_i, Y_i) | i = 1, \dots, n\}$ to obtain $\widehat{\delta}(x, y; d_x, d_y)$. Simar and Vanhems (2012) describe the way to proceed for the FDH estimator of Ψ , including their robust versions (order- m and order- α). They also provide all the statistical properties of the resulting estimators. Daraio and Simar (2014) gives all the practical details for computing $\widehat{\delta}$, in particular when some of the elements of d_x and/or of d_y are set to zero.

If we assume the convexity of Ψ we can rather employ the DEA estimator of $\delta(x, y; d_x, d_y)$ as follows:

$$\widehat{\delta}(x, y; d_x, d_y) = \sup\{\beta > 0 | (x - \beta d_x, y + \beta d_y) \in \widehat{\Psi}_{DEA}\}, \quad (2.3)$$

¹Since we will consider DDF, which are additive, we note that the inputs and/or the outputs can take negative values, see e.g. Kerstens and Van de Woestyne (2011).

where $\widehat{\Psi}_{DEA}$ is the well known free disposal and convex hull of \mathcal{X} . For instance, if the Variable Returns to Scale (DEA-VRS) version is used we have :

$$\widehat{\Psi}_{DEA} = \{(x, y) \in \mathbb{R}^{p+q} | y \leq \sum_{i=1}^n \lambda_i Y_i, x \geq \sum_{i=1}^n \lambda_i X_i, \text{ for some } \lambda_i \geq 0 \text{ with } \sum_{i=1}^n \lambda_i = 1\}, \quad (2.4)$$

where a version of the latter admits Constant Returns to Scale (DEA-CRS) if we delete the last constraint $\sum_{i=1}^n \lambda_i = 1$. Simar, Vanhems and Wilson (2012) derive the statistical properties of this estimator. To save space we limit our presentation to the DEA estimator, but once a direction vector is chosen, any nonparametric estimator can be used, including the FDH, their conditional versions and/or the robust partial order efficiency measures (see Simar and Wilson, 2014 for a recent survey on the statistical properties of nonparametric estimators of frontiers, including all the estimators listed before).

2.2 Selection of a Data-Driven Direction: the basic idea

Suppose we have to choose a direction in the input space $x \in \mathbb{R}^p$. Choosing a direction is equivalent to choosing an angle, for instance, if $X_i > 0$, the individual specific direction $d_{X_i} = X_i$ corresponds to the angle defining the input mix X_i . Mathematically, a direction x in the input space \mathbb{R}^p is described by its angles in polar coordinates (see the Appendix A on polar coordinates in multidimensional spaces). Denote by $r = \|x\|$ the modulus and $\theta = (\theta_1, \dots, \theta_{p-1})$ with $-\pi/2 < \theta_j \leq \pi/2$ for $j = 1, \dots, p-2$ and $-\pi < \theta_{p-1} \leq \pi$ the angles defining the direction x . By using the notations introduced in Appendix A for the traditional “à la Farrell” specific radial direction, we have $d_{X_i} = \phi(r_i, \theta_i)$ (or $d_{X_i} = \phi(1, \theta_i)$ if we want to use normalized directions).

A commonly used direction in the literature is the empirical *egalitarian* direction (i.e. an egalitarian direction based on the data) which consists in defining $d_x = \bar{X} = n^{-1} \sum_{i=1}^n X_i$ (which may be normalized afterwards, if needed). This approach will give more weights to angles corresponding to observed units having larger values of X_i , either because they are big units or maybe because they are very input-inefficient. We suggest instead a more objective average of directions given by the averages of the angles $\bar{\theta} = n^{-1} \sum_{i=1}^n \theta_i$. In this case, the normalization is left at our choice. We may choose $d_{X_i} = \phi(1, \bar{\theta})$ if we want normalized directions or $d_{X_i} = \phi(r_i, \bar{\theta})$ if a direction rescaled in original units is preferred. In this case, we observe that the resulting direction is obtained by a rotation of X_i to reach the average angle $\bar{\theta}$. Note that this approach based on averaging the angles is identical to the classical mean of the inputs \bar{X} if and only if all the original data X_i have the same length, which is of course quite improbable. Working with polar coordinates makes more clear the consequences

of this choice. To the best of our knowledge, this fact was never recognized in the existing literature.

Now, our idea is to consider cases where imposing the same direction (the same angle) to all the units being evaluated might be unreasonable due to contextual exogenous factors that may influence the choice of a particular input mix by the DMUs. The idea of taking a context-dependent direction is attractive. We look for a local direction that accounts for possible heterogeneity measured by some exogenous contextual factors, that we denote by W . Note that these variables could also be some inputs and/or outputs not under the control of the manager (see our comment above). The idea of our approach is to find a kind of local average of directional vectors where “local” is measured (or approximated) by these contextual factors. By applying our approach it may turn out, in extreme cases, when these factors have no effect on the production process, that the egalitarian average direction (that is the same direction for all the DMUs) will be reasonable; or, on the contrary, on the other extreme, we may find that these factors are so important that they justify the use of individual specific directions (that is a different direction for each DMU). What is appealing is that this choice will be the result of our data-driven technique. It will be, in fact, our approach that will tell us where we are between these two extreme cases.

In this section, we present the concept in a simplified case and for a direction in the input space. In Appendix B we give all the technical details for more general cases and the references for the statistical properties.

What we need is a regression model for finding $\mathbb{E}(\theta|W = w)$. Undoubtedly, to avoid too restrictive assumptions in the model, we will use a nonparametric regression framework. Suppose for simplicity that θ is univariate (the case $x \in \mathbb{R}^2$) and W is univariate, we will use the regression model:

$$\theta = m(W) + \varepsilon, \tag{2.5}$$

where the random angle ε has zero mean, finite dispersion and is independent of W , so that $\mathbb{E}(\theta|W = w) = m(w)$. In Appendix B we explain how to estimate $m(\cdot)$ from a sample of n observations (w_i, θ_i) , $i = 1, \dots, n$ including the case of multivariate θ and W . Roughly speaking, we could say that the resulting $\hat{m}(w)$ at any given value w of the contextual factors, is a weighted average of the angles θ_i corresponding to observations i such that W_i is not too far from w . This neighborhood of w is identified by a window width called the *bandwidth* and denoted by h and by a density function called the *kernel* function (like, e.g., a normal density centered on w and with standard deviation h) which is used to determine the weights in the average to be given at the angle θ_i , $i = 1, \dots, n$ as a function of the differences $(W_i - w)$. It is very similar to regular nonparametric regressions (see Li and Racine, 2007 or Pagan and Ullah, 1999), but with some changes due to the angular nature of the dependent

variable and so the notion of distance has to be adapted to this case. The details are given in Appendix B.

It is important to note the great flexibility of our approach: some elements of the direction vector d_x could be fixed for strategic reasons, or set equal to zero, then the angular regression above could be used only for the input mix that are not fixed.

Hence, we end-up with a very natural concept of “local” direction, determined by the nearest neighborhoods DMUs, where the proximity is measured according to the variable W . The choice of the kernel function is not so important in contrast with the choice of the bandwidth. Hopefully the theory of nonparametric regressions applies and, as explained in Appendix B, optimal bandwidths are selected by data-driven techniques (Least-squares cross validation). As recalled above, a great advantage of our approach is its flexibility and its ability to cover the two extreme cases mentioned above for choosing the direction. It has been shown in Hall, Li and Racine (2007) that if a variable W is irrelevant and has no effect on the dependent variable, $h \rightarrow \infty$ and so the “local” average is done over a window of infinite width, i.e. over all the data points and we are back to the overall mean (egalitarian case) as the resulting direction: $\hat{m}(w) = n^{-1} \sum_{i=1}^n \theta_i$ for all w . If the effect of W is very strong, the value of the bandwidth will be small, at the limit $h \rightarrow 0$, and for estimating $m(W_i)$, we will take an average using only the single angle θ_i , so $\hat{m}(W_i) = \theta_i$ for all $i = 1, \dots, n$ going back to the individual-specific direction case.

The same reasoning applies also for the output space $y \in \mathbb{R}^q$, with the same and/or other contextual factors as well as to more general input-output space analyses. The practical implementation of this approach in our frontier estimation context is described in Section 4.

3 Benchmarking

The idea of benchmarking is very common in the literature of productivity analysis (see e.g. Thanassoulis et al., 2008 and the references cited therein). It is indeed one of the key outcomes of efficiency analysis the identification of targets and efficient peers to benchmark the DMUs under evaluation. Benchmarking activity is also important to support strategic decision making (see e.g. Dyson, 2004 for a link with swot, -strengths, weaknesses, opportunities and threats-, analysis and resource-based planning). In most of the existing works, the target is the “virtual” unit obtained by the radial projection of the unit under evaluation on the DEA frontier, along the ray defined by their input (or output) mix. Traditionally, efficient peers are usually “best practice” units lying on the DEA efficient frontier that are in a sense not too different in terms of their inputs and outputs compositions from the unit being evaluated. They are identified by the units defining the efficient facet of the DEA fron-

tier where the benchmarked unit is radially projected. Often, radar plots are then provided where the inputs and the outputs of the benchmarked unit are compared to the inputs and the outputs of the target and of the observed efficient peers. The idea is that the comparison is done with existing units “not too far” from the benchmarked unit in terms of inputs and outputs mixes but being “efficient” (see Thanassoulis et al. pp 357–359 for some examples).

When directional distance functions are used the target is then defined as the virtual unit obtained by the projection of the evaluated unit to the DEA frontier, along the chosen direction. Again, this point belongs to one of the facets of the DEA efficient frontier which could be chosen for identifying the efficient peers against which the unit could be benchmarked. However, these units may not be the most appropriate ones, in particular if some contextual, specific external factors (like the W introduced above) may influence the chosen input and/or output mixes. In such a case, the procedure introduced in the preceding section offers a nice way to identify better efficient units to serve as benchmark for an inefficient unit.

Indeed, we can decide to consider in the benchmarking process of a unit facing the value w for these contextual factors, only units facing “similar” conditions. These units have already been identified by our approach. This was done in the phase determining the local average direction for the evaluated unit: they are those units such that $|W_i - w| \leq ch$, where $c \leq 1$ is a constant tuning the chosen neighborhood.² In our application below we choose $c = 1/2$ for getting units with values W_i within a range of one bandwidth around the current value w and the value of h is obtained by the data driven method described in the Appendix B, see equation (B.8). Among these points, it is easy to identify those lying on the DEA frontier (having values $\widehat{\delta}(X_i, Y_i; d_{X_i}, d_{Y_i}) = 0$). Finally, among the remaining points we select the benchmarked units as being the, say 3 to 5, nearest neighbors in the $p + q$ full dimensional space of inputs and outputs, where for variables such that $d = 0$ (inactive variables in the optimization), original euclidean units are used, but for those with positive directional elements $d > 0$ (active variables in the optimization) only the mix is important, so we use normalized units (see step [5.2] of the practical algorithm in the next section).

Finally, for visual inspection, as done in the literature, we can show radar plots representing the input and output values of the evaluated DMU, of the target virtual unit and of one of the few benchmark units selected above.

There are two main differences of our approach with the traditional benchmarking approach. First of all, in our approach, the radar plots will never use as benchmark units those units facing too different external-environmental conditions. Secondly, the benchmarking

²When W is multivariate, this inequality has to be understood component wise: $|W_i^j - w^j| \leq ch^j$ where the relative notation is introduced in Appendix B.

within our approach is fully data-driven, based on the optimal selected bandwidths, and is automatically done without having to carry out an ad hoc clustering exercise.

In Section 4 we give the practical details on how to implement our procedure which is illustrated with some examples in Section 5.

4 Practical Methodological Steps

In this section we describe the main steps of our approach. To avoid complexity in notation and presentation, we will refer hereafter to an input orientation case where the direction vectors in the input space is $d_x \geq 0$ and for the output space we have $d_y = 0$. As above $x \in \mathbb{R}^p$, $y \in \mathbb{R}^q$ and the heterogeneity conditions that may influence the rays (input and or output mixes) of DMUs are denoted $W \in \mathbb{R}^d$. The approach can easily be adapted when the directions are “active” in the output case only ($d_y \geq 0$ and $d_x = 0$) but also when both input and output directions have “active” components. The main phases are as follows.

- [1] We transform each observed p -dimensional input X_i in polar coordinates (r_i, θ_i) , for $i = 1, \dots, n$, where $r_i > 0$ and $\theta_i = (\theta_i^1, \dots, \theta_i^{p-1})$ (see Appendix A for details).
 - [1.1] Eventually, look at the histograms of θ^j , $j = 1, \dots, p - 1$ over the n data points to have a descriptive idea of the heterogeneity of the input mixes (note that this can be done in radians or in degrees). Note also that for $p = 3$, most softwares produce 3D-Histograms for the 2 angles.
- [2] Perform the polar nonparametric regression for each component θ^j on W to estimate $\mathbb{E}(\theta^j|W)$. For this we use for each regression j , $j = 1, \dots, p - 1$ the set of data (θ_i^j, W_i) , $i = 1, \dots, n$.
 - [2.1] The bandwidth selection for each regression is computed by cross-validation (see (B.8)). Note that here we use rather Local-Constant Least-Squares estimates to ensure that the fitted values will stay in the same quadrants as the original values, which could not be the case when using Local-Linear Least-Squares.
 - [2.2] By applying (B.9) and (B.10) we obtain (B.5) and hence we obtain for each angle $j = 1, \dots, p - 1$ and for each data point, the local average estimator

$$\hat{\theta}_i^j = \hat{\mathbb{E}}[\theta^j|W_i] = \hat{m}(W_i). \quad (4.1)$$

Note that a natural measure for the Quality of the Fit (QF) can be given by the average of the angular distances between $\hat{\theta}_i^j$ and the original θ_i^j :

$$QF = \frac{1}{n} \sum_{i=1}^n \left[1 - \cos(\hat{\theta}_i^j - \theta_i^j) \right], \quad (4.2)$$

where a value of 0 indicates a perfect fit.³

[2.3] We now have for each $i = 1, \dots, n$ the angles $\widehat{\theta}_i = (\widehat{\theta}_i^1, \dots, \widehat{\theta}_i^{p-1})$

[2.4] Optionally, we can provide some partial plots of the resulting angles $\widehat{\theta}_i$ as a function of W_i , to investigate the effect of each component of W on the resulting directions. For instance if W is bivariate, we can provide 3D plots of the surface determined by an angle $\widehat{\theta}^j$ as a function of the two components of W .

[3] From the polar coordinates $(r_i, \widehat{\theta}_i)$ we go back to cartesian coordinates giving the directional vector d_{X_i} so we have $d_{X_i} = \phi(r_i, \widehat{\theta}_i)$ (or $d_{X_i} = \phi(1, \widehat{\theta}_i)$ if we prefer to work with normalized directional vectors).

[4] Compute the DDFs for each DMUs. In our illustration we chose to do it with the 3 following options for the distance vectors:

i) individual specific distance $d_{X_i} = X_i = \phi(r_i, \theta_i)$;

ii) our data-driven local averages $d_{X_i} = \phi(r_i, \widehat{\theta}_i)$; and

iii) the egalitarian approach where $d_{X_i} = \phi(r_i, \bar{\theta})$ corresponding to the angle $\bar{\theta} = n^{-1} \sum_{i=1}^n \theta_i$.

Undoubtedly, for the 3 cases we can use their normalized version if desired.

[4.1] Clearly our approach will give results in-between the two extreme cases. As a matter of fact, this will depend on the data through the data-driven selected bandwidth. If h is large, $\widehat{\theta}_i \approx \bar{\theta}$ and if h is very small, $\widehat{\theta}_i \approx \theta_i$. The flexibility of our approach, which uses product kernels, allow us to analyze the influence of each W^j , component wise.

[4.2] Undoubtedly, subsequently the traditional efficiency analysis is applied for the 3 cases; that is, efficiency measures $\widehat{\delta}(X_i, Y_i; d_x, 0)$ for each DMU as well as the gaps in original units of the inputs (or in percentage of the inputs), given by $\widehat{\delta}(X_i, Y_i; d_x, 0) \times d_x$, to reach the efficient frontier, are calculated.

[5] Benchmarking analysis. For any particular DMU of interest, characterized by the data value (x, y, w) (typically any observation (X_i, Y_i, W_i) in the original sample), we can apply the benchmarking procedure described in Section 3. The procedure may be summarized as follows:

³In nonparametric regressions, for assessing the goodness of fit, Racine (2008) motivates also the use of $R^2 = \left[\sum_{i=1}^n (\theta_i - \bar{\theta})(\widehat{\theta}_i - \bar{\theta}) \right]^2 / \sum_{i=1}^n (\theta_i - \bar{\theta})^2 \sum_{i=1}^n (\widehat{\theta}_i - \bar{\theta})^2$. We have $R^2 \in [0, 1]$ with the value 1 for perfect fit and value 0 for no predictive power of the model.

- [5.1] Identify the units in a W -neighbor of w by selecting units i with $|W_i^j - w^j| \leq 0.5h^j$, $j = 1, \dots, d$. Suppose we have n_w such units. Among these n_w units select the efficient ones lying on the DEA frontier (having $\widehat{\delta}(X_i, Y_i; d_x, 0) = 0$). Now we have n_w^∂ such points.
- [5.2] Now we will select the nearest neighbors of the ray x among the n_w^∂ selected points. By doing this we also take into account for the output levels y and Y_i .
- [5.2.1] For doing this we project the inputs x and X_i of the n_w^∂ points on a sphere of radius one. This defines x_s and $X_{i,s}$. It is important to note that we are doing this because in the input oriented case, only the mix of the inputs are important and not the values.
- [5.2.2] Compute the Euclidean (or Mahalanobis, if preferred) distances between (x_s, y) and the n_w^∂ points $(X_{i,s}, Y_i)$, and select the n_B nearest points according to this metric (in practice n_B is limited to a few points).
- [5.3] Make a radar plot of (x, y) against the efficient *target* (its projection on the efficient frontier) and the n_B units selected for the benchmarking. Usually we will provide n_B radar plots to facilitate the analysis. Also we chose to standardize at 1, the unit to be evaluated, (x, y) , to facilitate the interpretation in terms of percentage of increase or decrease of the $p + q$ inputs and outputs to reach the target or the selected benchmarked unit.

In the next section we illustrate this methodology with a simulated sample and two empirical examples with real data.

5 Applications

5.1 Simulated example

We first present some results in a simulated sample with two inputs X , two outputs Y and two external factors W . We select an output orientation, so we fix $d_X = 0$ and we will illustrate the choice $d_{Y_i} = \phi(1, \widehat{\theta}_i)$ obtained by the nonparametric angular regression (local average) and compare afterwards the obtained results with the two extreme choices, individual specific $d_{Y_i} = Y_i/||Y_i|| = \phi(1, \theta_i)$ and the egalitarian $d_{Y_i} = \phi(1, \bar{\theta})$ (note that we normalize the directions). To simplify the notation we denote by W_1 and W_2 the two components of W . An external factor (W_1) will be influential in the determination of the output mixes while the other (W_2) will be completely independent with respect to the production process. We will consider the case of a sample of size $n = 100$ which is not big with respect to the dimension

of the problem considered because the idea is to see if we are able to recover by our approach the DGP used in our simulated scenario.

The simulation scenario can be described as follows.

We simulate random output mixes by generating an angle $\theta \in [0, \pi/2]$ which is dependent on the variable W_1 through a logistic model plus noise. We define

$$\theta = (\pi/2) \frac{\exp(W_1)}{1 + \exp(W_1)} + \varepsilon, \quad (5.1)$$

where $W_1 \sim \text{Beta}(3, 3)$ rescaled so that $W_1 \in (a, b)$ with a and b fixed such that the factor $\exp(W_1)/(1 + \exp(W_1)) \in (0.025, 0.975)$ and ε is a censored $N(0, 0.2^2)$, censored such that $\theta \in [0, \pi/2]$.⁴ We have to truncate the normal to avoid generated values of θ outside $[0, \pi/2]$ due to the random noise. To fix the ideas, we display in Figure 1 the histogram of the simulated output mixes. Figure 2 (right panel) display the true logistic regression and show such a simulation for $n = 100$ where it appears clearly that W_1 has a “strong” influence on the output mix (high heterogeneity). Then we simulate value of W_2 independently from the production process by means of the uniform $W_2 \sim \text{Unif}(1, 10)$.

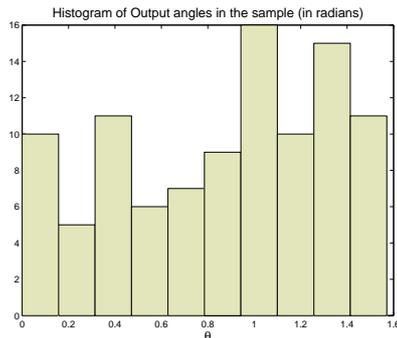


Figure 1: *Histogram of the observed output mixes (angles) $\theta = (\theta_1, \dots, \theta_n)$ in the Simulated Example 1, with $n = 100$.*

The angular nonparametric regression procedure selected the bandwidths $h_1 = 1.0359$, roughly 1/6 of the range of W_1 , and $h_2 = 2.9768$, roughly 1/3 of the range of W_2 , indicating clearly less influence for the latter in the variations of the angle. The fit appears very good as displayed in Figure 2 (left panel): we are able to recover the logistic influence of W_1 and the “no-influence” of W_2 . The right panel of Figure 2 is a marginal view of the left panel from an orthogonal viewpoint to the axis W_1 . The red squares are the obtained fit at the observed $W_{1,i}$ (they are very near the true logistic function) whereas the blue triangles are

⁴It is easy to show that this is achieved by choosing $a = \log(0.025/(1 - 0.025))$ and $b = \log(0.975/(1 - 0.975))$.

the observations. The good quality of the fit can be measured by $R^2 = 0.8649$ and a very low angular distance, defined in (4.2) (see also Appendix B) with value 0.0142, between the observations and the fits.

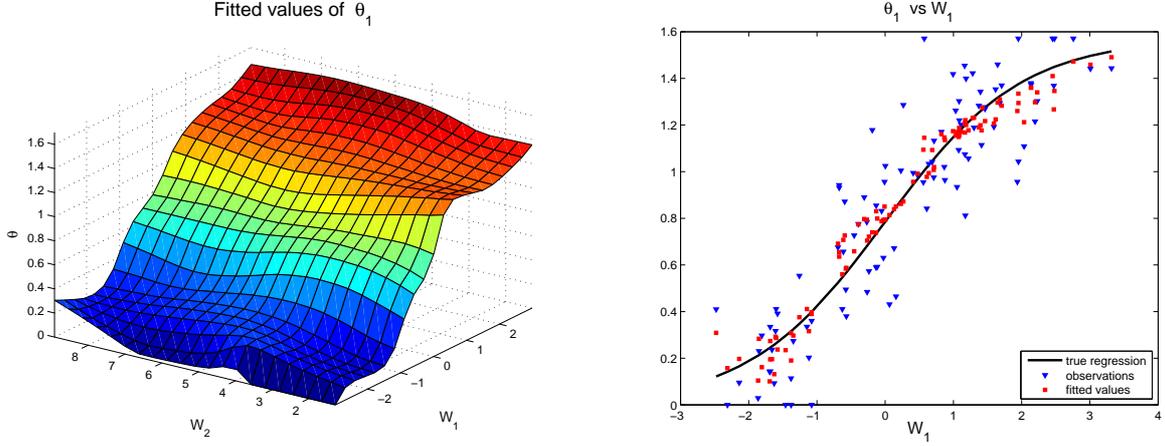


Figure 2: *Fitted values of the angle θ by the nonparametric regression of θ on $W = (W_1, W_2)$, in Simulated Example 1, with $n = 100$.*

For simulating the production process, we fix the modulus of production frontier as being of a Cobb-Douglas form $R = X_1^{0.3} X_2^{0.4}$, where both inputs are independently generated by the uniform $\text{Unif}(10, 20)$. Finally the two outputs generated on the frontier are determined by the simulated angle θ , where we chose the formulation in (A.2), i.e.

$$\begin{cases} Y_1 = R \sin \theta \times \exp(-U) \\ Y_2 = R \cos \theta \times \exp(-U), \end{cases}$$

where the last factor accounts for radial inefficiencies ($U \geq 0$). We have chosen $U \sim \text{Expo}(1/3)$ which implies $\mathbb{E}(\exp(-U)) = 0.75$ as average level of Farrell Debreu output inefficiency.

The resulting directional distances have been computed using the 3 methods proposed above (individual specific direction; our W -local averages and the global average direction or egalitarian common direction). In Table 1 we report the averages DDF for the $n = 100$ units and the averages of the gaps to reach the efficient frontier in the two output components. As expected by our scenario, due to the great heterogeneity introduced by a strong dependence between the angles θ and the variable W_1 , we obtain, by our data-driven technique a solution which is not too different from the one where a “full” heterogeneity is allowed, i.e. using individual specific directions. Note the great difference between the gaps when using the egalitarian distance and the two other approaches. Our data-driven procedure warns the research analyst that here egalitarian distance is inappropriate due to the impact of the

external specific factors (here W_1 on the selection of the output mixes. In the next simulated case we will have an opposite conclusion.

	$d_{Y_i} = \phi(1, \theta_i)$	$d_{Y_i}(W_i) = \phi(1, \hat{\theta}_i)$	$d_{Y_i} = \phi(1, \bar{\theta})$
$\overline{\text{DDF}}$	0.9810	1.0054	1.1053
$\overline{\text{Gap}}_{Y_1}$	0.7006	0.7321	0.8410
$\overline{\text{Gap}}_{Y_2}$	0.5414	0.5830	0.7172

Table 1: *Simulated Example 1 with $n = 100$. Averages DDFs, and average of the gaps in the two outputs for reaching the efficient frontier. The transformation $\phi(1, \cdot)$ is the one defined in Appendix A.*

We now modify slightly the scenario above by letting both components of W be independent of the production process. So these two components have no influence on the determination of the output mixes. We simulate now the angles by $\theta \sim \text{Unif}(0, \pi/2)$. Except for that, all the elements of the scenario remain the same.

Our data-driven procedure provides the bandwidths $h_1 = 6.2809$ and $h_2 = 8.7472$, both being of the same order than the range of W , indicating clearly no influence for the latter in the variations of the angle. The goodness of fit is now very bad giving $R^2 = 0.0146$ and an angular distance equal to 0.0885. The result of the fit of the angles by the nonparametric regression on W is displayed in Figure 3, with the two marginal views. This confirm that our procedure automatically detect the “no-influence” of W on θ , as it should. Since there is no heterogeneity due to W , an egalitarian approach seems more reasonable and this is not far from what we obtain by looking at Table 2. The mean of efficiencies and the averages of the gaps for the two outputs derived from our procedure are quite similar to the ones obtained by taking the egalitarian direction $d_y = \phi(1, \bar{\theta})$. The nice feature (aspect) is that we let the analysis (i.e. the data) to derive the chosen directions, and it is not decided *a priori* (data driven approach).

	$d_{Y_i} = \phi(1, \theta_i)$	$d_{Y_i}(W_i) = \phi(1, \hat{\theta}_i)$	$d_{Y_i} = \phi(1, \bar{\theta})$
$\overline{\text{DDF}}$	1.2662	1.3914	1.3910
$\overline{\text{Gap}}_{Y_1}$	0.7737	0.9199	0.9224
$\overline{\text{Gap}}_{Y_2}$	0.8697	1.0438	1.0412

Table 2: *Simulated example 2 with $n = 100$. Averages DDFs, and average of the gaps in the two outputs for reaching the efficient frontier.*

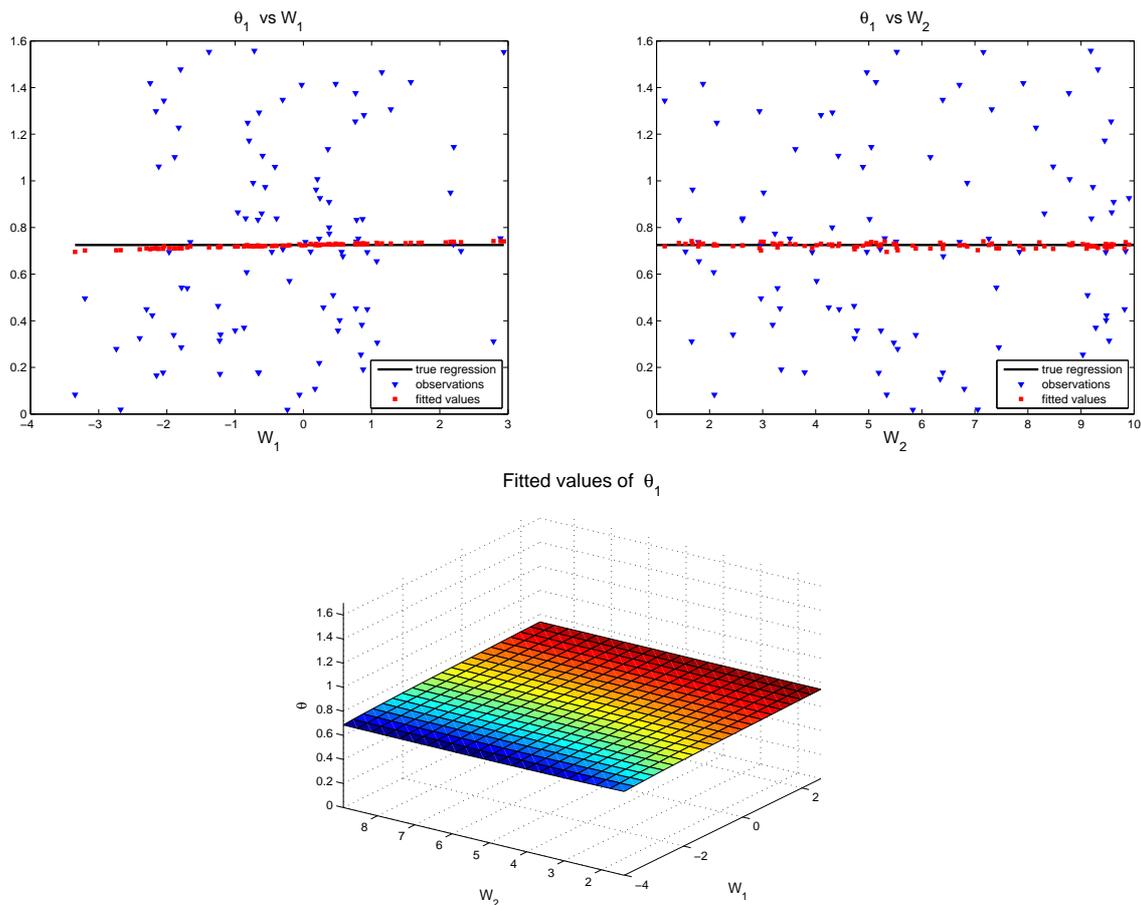


Figure 3: *Fitted values of the angle θ by the nonparametric regression of θ on $W = (W_1, W_2)$, in Simulated Example 2 with $n = 100$.*

5.2 Illustration with real data on education

We illustrate the usefulness of our approach as a tool for selecting an objective (data driven) direction by analyzing a “production process” with the popular data set from Charnes, Cooper and Rhodes (1981) where the performance of 70 schools is analyzed, 49 of them having benefited from Program Follow Through (PFT) and 21 called Non-Follow Through (NFT). The paper gives the data on 3 outputs (achievement of students on “Reading Score” (READ), ”Mathematics Scores” (MATH) and a measure of “Self-Esteem” (SELF)). The only real input is the number of teachers. The paper uses also 4 variables as inputs. In fact these variables are not inputs under the control of the School but rather describe 4 characteristics of the family (in units of 100 students). These are the “Education Level of the Mother” (ELM), an index for the “Highest Occupation of a family member” (HOFM), a “Parental Visit Index” (PVI) and a “Parent Counseling Index” (PCI). These four latter

variables are highly correlated so for ease of presentation we keep only ELM and PVI, the lowest correlated pair (0.7936). Note that HOFM and PCI have a correlation above 0.98 with PVI, so they do not bring useful additional information.

We will use an output orientation to see how the schools are performing in using their only input, the number of teachers. The variables describing the profile of the family (level of education and involvement in the following-through of the children, as measured by ELM and PVI) may be viewed in our approach as describing contextual specific environment (profile of the family) and we analyze if they have any influence on the output mix of the school, described here by the two angles. If yes, and if the influence is large, this could justify the use of a school-specific direction for assessing and comparing the efficiencies (as done in the original study of Charnes et al., 1981), but if not and if no particular direction is desired, a natural direction might be a common-egalitarian direction for all the schools. Our approach will determine where we are between these two extremes.

In our analysis we eliminate two schools (unit # 44 and unit #59) because they are really outliers (as detected in many studies, including Wilson, 1993 and Simar, 2003). In fact these units have values above the mean plus three times the standard deviation (for both ELM and PVI for #59 and for PVI for #44), making not available a reasonable non-parametric estimate of the regression below because these two points are too isolated in the W space. So we end up with 68 units.

The distribution of the two angles defining the output mixes is described in degrees in Figure 4. The analysis is interesting because the figure indicates that there are no real big differences in the output mixes (this fact was already noticed in preceding studies and confirmed by the high correlation between the three outputs), but still there is some dispersion and we can pursue our analysis. The angle θ_1 ranges roughly between 25 and 42 degrees and the angles θ_2 between 43 and 63 degrees. Looking at Figure 9 in Appendix A, we see that *increasing* values of θ_1 indicates a school that “select” as individual target to increase proportionally more the READ variable Y_1 relative to the two others. A *decreasing* value of θ_2 indicates a school targeting a proportional increase of more the variable SELF (Y_3) than MATH (Y_2).

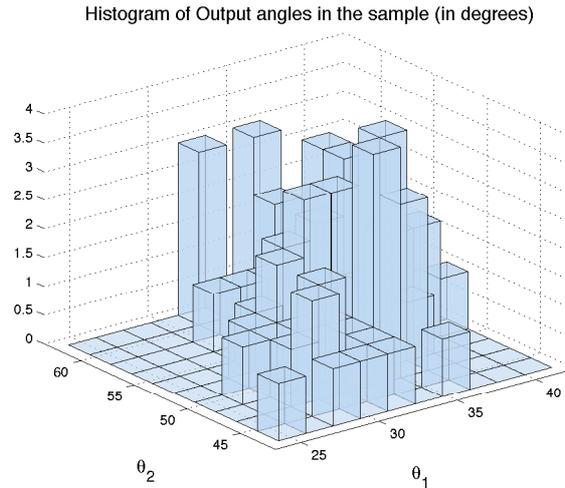


Figure 4: *Histogram of the observed output mixes (angles) $\theta = (\theta_1, \dots, \theta_n)$ in the CCR-School Example, with $n = 68$.*

The results of the nonparametric polar regression of these two angles on the variables W_1 (ELM) and W_2 (PVI) are displayed in Figure 5. We see that as expected from the beginning (small variations in the output mixes), the two surfaces are rather flat, indicating a low level of influence of W on the output mixes. Still, we see some effect (with interaction) of both variables in the following sense. For schools where the level of education of the mother is higher, the visiting parental effect is null; on the contrary, for schools where the level of education of the mother is low there is some negative effect of IPV on both angles. We see also that only for schools where the IPV index is high (many parental visits), the role of the level of education is more important and positive. Due to the interpretation of the angles done above we could say (being careful because the analysis relies on a small sample) that higher ELM will slightly increase the weight of the reading scores in the school, but more for school where the IPV is relevant (important). On the other hand, higher IPV values increase the weight of the self-esteem index (relative to the mathematics and reading scores) but mainly for schools where the education level of the mother is the lower. This kind of analysis can have its own interest, even before speaking about performance of the schools themselves.

The technical results of these two nonparametric regressions are quite similar: the two bandwidths in both regression are rather high but still significant (of the order of 1/4 of the range of W_1 and 1/3 of the range for W_2). The R^2 is 0.36 for $\hat{\theta}_1$ (angular risk 0.0012) and 0.23 for $\hat{\theta}_2$ (angular risk 0.0032). So a relatively good fit in angular distance measure.

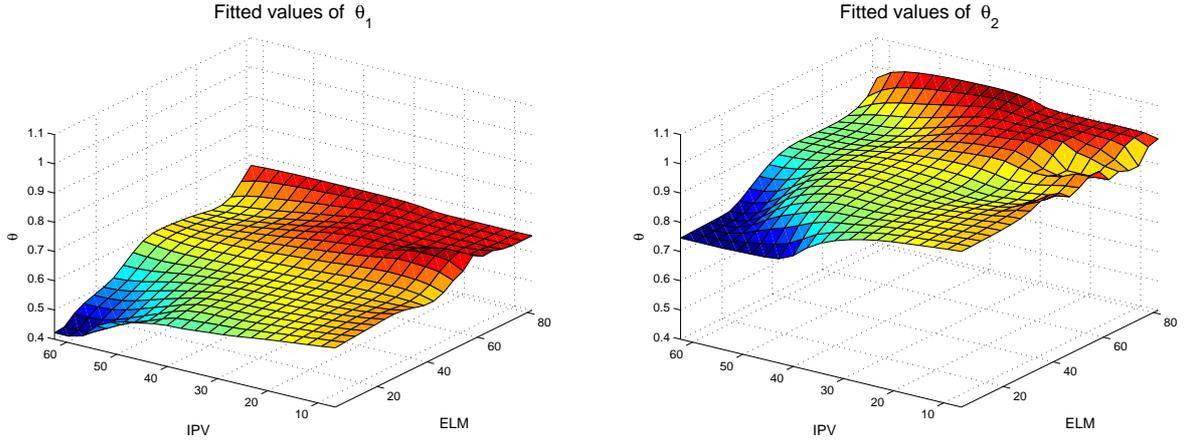


Figure 5: *Fitted values of the angle θ by the nonparametric regression of θ on $W = (W_1, W_2)$, in the CCR-School example with $n = 68$.*

In our illustration here, the individual efficiency scores of each school are not of central interest, but we provide the averages below according to the chosen direction (remember that efficient units have an efficiency score equal to zero). We expect in this particular example very small differences according the 3 directions (individual specific, W -local averages or global common average) but still we can observe some differences with our data driven approach being this time very similar to the egalitarian (global average direction) which is not a big surprise due to the “flatness” of both surfaces in Figure 5. The results are given in Table 3. With the model chosen here (one input and 3 outputs), there might be some significant differences between the two groups of schools. Formal tests as those suggested by Simar and Wilson (2011) using the bootstrap or by Kneip, Simar and Wilson (2013) using asymptotic results could be used to check the significancy. This is not the focus here.

	$d_{Y_i} = \phi(1, \theta_i)$	$d_{Y_i}(W_i) = \phi(1, \hat{\theta}_i)$	$d_{Y_i} = \phi(1, \bar{\theta})$
DDF	20.9507	22.0725	22.0725
DDF _{PFT}	17.0682	18.1488	18.1488
std	(2.2046)	(2.3952)	(2.3952)
DDF _{NFT}	30.2687	31.4894	31.4894
std	(3.6993)	(3.8283)	(3.8283)

Table 3: *CCR-example. Averages DDFs, and average for the two groups of schools, with standard deviations of within group averages.*

To save place we do not give more detailed results (like individual measures or evaluation of the gaps), because the main interest of this illustration is to see the usefulness of the

analysis of the angles as a function of the contextual variables, here the profile of the families in each school. Here also benchmarking is not our focus, although in a real application, as illustrated in the next example, this might be of interest.

5.3 Empirical application in Banking

The original data set contains three inputs (purchased funds, core deposits and labor) and four outputs (consumer loans, business loans, real estate loans, and securities held) for banks. Aly et al.1990 considered, among others, two continuous environmental factors, the size of the banks, and a measure of the diversity of the services proposed by the banks (see Aly. et al.,1990, for details). This data set has also been used in Simar and Wilson (2007) where the measure of the size of the banks is given by the log of the total assets, rather than the total deposits as in Aly et al. We will illustrate our procedure with the subsample of 322 banks used in Simar and Wilson (2007). We will use the input orientation (that is $d_y = 0$ and $d_x \geq 0$). Hence, we will have two angles defining the input mixes and we will use the two contextual variables W_1 (SIZE) and W_2 (DIVERSITY) to see if they have some influence on the determination of the input mixes, and so justifying, at some level, individual specific directions for the DMUs. We will also illustrate our benchmarking procedure for one particular inefficient bank.

In our sample, the distribution of the input mixes is represented in Figure 6. We see from the start that the angle θ_2 has almost no variations (most of the data points have an input mix near and below 90 degrees. This is due of the chosen units for X_3 , number of workers and X_2 , core deposit in dollars. However, we see a greater dispersion of the angle θ_1 , between 0 and 60 degrees, with most of the data with an angle below 40 degrees. Here an increasing angle corresponds to a more important weight of purchased funds mainly relative to core deposit (due to the low weight of labor in the available units). In this application, due to the huge differences between the units of the input components, we will use vector of distances in the units of the inputs to obtain a measure that is free of the chosen units.

With no surprise, the nonparametric regression of θ_2 on (W_1, W_2) is almost flat as shown in Figure 7 (the bandwidths are $h_1 = 1.11$ and $h_2 = 1.12$ with $R^2 = 0.05$ and the angular risk is below 0.00005). However we find some structure in the same figure, left panel, where the expectation of θ_1 is changing with the variables (W_1, W_2) (here $h_1 = 2.05$ and $h_2 = 0.28$ with $R^2 = 0.27$ and angular risk is very low 0.0102). Here we see that banks with higher size and lower diversity index, have on the average higher values of θ_1 in their input mix, which can be interpreted here as proportionally more purchased funds than core deposit (again the labor has little weight in determining the angles). The effect of diversity is more important for bigger banks.

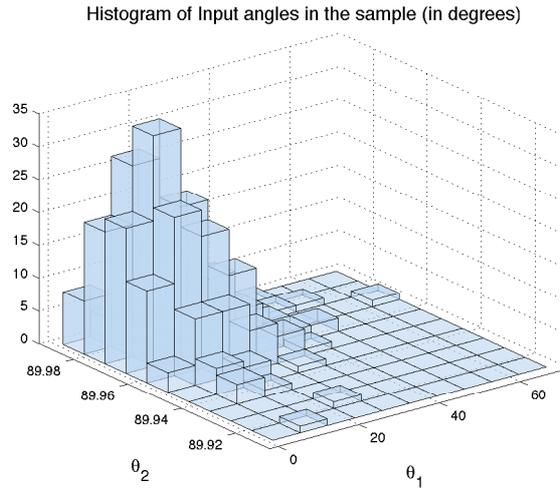


Figure 6: *Example on Banks: histogram of the input mixes in the sample of 322 banks.*

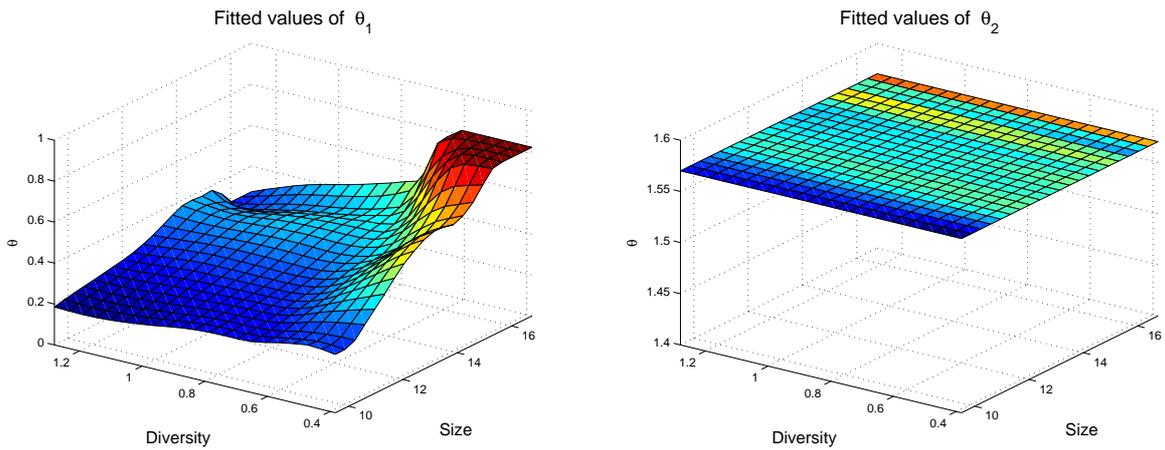


Figure 7: *Fitted values of the angle θ for the example on banks.*

The efficiency analysis with the three approaches (individual specific, local average and egalitarian) are reported in Table 4 for 20 units drawn at random. The mean at the bottom row is the mean over the full sample of 322 units. The differences between the 3 approaches seem to be not so important (remember that most banks share the same input mix, angle θ_2) but remember that these DDF are applied to directional distances in the same units as the original inputs. So even small differences of δ could have important consequences on the distance to the efficient frontier (as measured in gaps). This is shown by Table 5 giving the gaps in each inputs, in original units, for the same 20 units. We see here that changing the orientation matters, when we measure the effort to reach the efficient frontier (see e.g. unit #56). Table 6 gives the same gaps in percentage of the inputs.

Units	$d_{x_i} = \phi(r_i, \theta_i)$	$d_x(W_i) = \phi(r_i, \hat{\theta}_i)$	$d_x = \phi(r_i, \bar{\theta})$
204	0.0544	0.0563	0.0566
189	0.0000	0.0000	0.0000
127	0.0446	0.0403	0.0388
78	0.0736	0.0713	0.0696
299	0.0398	0.0362	0.0358
190	0.0000	0.0000	0.0000
82	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000
56	0.0364	0.0277	0.0331
301	0.0565	0.0544	0.0525
152	0.0000	0.0000	0.0000
148	0.0253	0.0225	0.0223
92	0.0000	0.0000	0.0000
173	0.0729	0.0754	0.0756
225	0.0014	0.0011	0.0011
321	0.0244	0.0258	0.0254
157	0.0300	0.0305	0.0306
68	0.0000	0.0000	0.0000
139	0.0104	0.0101	0.0102
155	0.0000	0.0000	0.0000
$mean_n$	0.0335	0.0328	0.0323

Table 4: *Example on banks: DDFs for 20 units.*

Units	$d_{x_i} = \phi(r_i, \theta_i)$			$d_x(W_i) = \phi(r_i, \hat{\theta}_i)$			$d_x = \phi(r_i, \bar{\theta})$		
	Gap $_{X_1}$	Gap $_{X_2}$	Gap $_{X_3}$	Gap $_{X_1}$	Gap $_{X_2}$	Gap $_{X_3}$	Gap $_{X_1}$	Gap $_{X_2}$	Gap $_{X_3}$
204	6283.10	10829.04	9.3030	3712.71	12411.30	6.9617	3516.82	12531.89	7.0932
189	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
127	166.05	1335.03	0.7576	275.45	1183.78	0.6906	316.34	1127.26	0.6380
78	707.71	3576.17	2.5033	834.93	3428.75	1.9101	930.98	3317.45	1.8777
299	237.43	1034.41	0.3980	200.88	944.02	0.5544	257.60	917.94	0.5196
190	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
82	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
2	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
56	591.99	2867.83	1.1290	1069.82	1952.19	1.0457	718.99	2562.07	1.4502
301	138.08	945.24	0.7913	190.19	899.60	0.5470	239.66	854.00	0.4834
152	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
148	163.04	1180.34	0.8085	273.81	1026.49	0.5722	284.06	1012.24	0.5729
92	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
173	5302.65	10881.98	5.6137	3549.95	12002.22	6.6818	3392.42	12088.59	6.8423
225	10.54	114.69	0.0698	22.54	92.21	0.0505	24.73	88.11	0.0499
321	781.93	1296.94	1.1239	370.85	1552.57	0.8548	426.06	1518.23	0.8593
157	3349.24	3514.01	2.1602	1436.93	4713.46	2.6148	1339.83	4774.37	2.7024
68	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
139	372.11	1475.69	0.7625	420.26	1412.62	0.7774	401.73	1431.51	0.8103
155	0.00	0.00	0.0000	0.00	0.00	0.0000	0.00	0.00	0.0000
$mean_n$	893.92	2807.85	1.7285	824.85	2796.33	1.5554	794.22	2830.13	1.6019

Table 5: *Example on banks: Gaps for 20 units.*

Units	$d_{x_i} = \phi(r_i, \theta_i)$			$d_x(W_i) = \phi(r_i, \hat{\theta}_i)$			$d_x = \phi(r_i, \bar{\theta})$		
	Gap $_{X_1}$	Gap $_{X_2}$	Gap $_{X_3}$	Gap $_{X_1}$	Gap $_{X_2}$	Gap $_{X_3}$	Gap $_{X_1}$	Gap $_{X_2}$	Gap $_{X_3}$
204	0.0544	0.0544	0.0544	0.0321	0.0624	0.0407	0.0305	0.0630	0.0415
189	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
127	0.0446	0.0446	0.0446	0.0739	0.0395	0.0406	0.0849	0.0376	0.0375
78	0.0736	0.0736	0.0736	0.0869	0.0706	0.0562	0.0969	0.0683	0.0552
299	0.0398	0.0398	0.0398	0.0337	0.0363	0.0554	0.0432	0.0353	0.0520
190	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
82	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
56	0.0364	0.0364	0.0364	0.0658	0.0248	0.0337	0.0442	0.0325	0.0468
301	0.0565	0.0565	0.0565	0.0779	0.0538	0.0391	0.0981	0.0511	0.0345
152	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
148	0.0253	0.0253	0.0253	0.0424	0.0220	0.0179	0.0440	0.0217	0.0179
92	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
173	0.0729	0.0729	0.0729	0.0488	0.0804	0.0868	0.0466	0.0810	0.0889
225	0.0014	0.0014	0.0014	0.0029	0.0011	0.0010	0.0032	0.0011	0.0010
321	0.0244	0.0244	0.0244	0.0116	0.0292	0.0186	0.0133	0.0286	0.0187
157	0.0300	0.0300	0.0300	0.0129	0.0402	0.0363	0.0120	0.0408	0.0375
68	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
139	0.0104	0.0104	0.0104	0.0118	0.0100	0.0106	0.0113	0.0101	0.0111
155	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$mean_n$	0.0335	0.0335	0.0335	0.0362	0.0332	0.0323	0.0392	0.0326	0.0319

Table 6: *Example on banks: Gaps for 20 units in percentages of inputs.*

For the benchmark analysis we selected the unit #7 only for illustrative purposes. There were 56 units with similar contextual values W and among these, there were 11 efficient units. Table 7 gives the list of the 4 nearest neighbors in the input-output space where the inputs are projected on the sphere of radius 1.

Units	DDFs	Euclid. dist.	Angul. dist.
7	0.0711	0	0
174	0.0000	4991.46	0.000046
46	0.0000	8787.05	0.000010
245	0.0000	9100.66	0.000145
219	0.0000	9415.60	0.000025

Table 7: *Selected efficient benchmark units for unit #7. The Euclidean distance between units is in \mathbb{R}^{p+q} , original outputs and inputs projected on the Sphere of radius 1. The angular distance is between the input mixes.*

The radar plots of the 4 nearest efficient banks selected by the procedure described above are given in Figure 8. The analysis of these pictures could help the manager to see where efforts could be put for improving its performance. The availability of several benchmark

units offers the possibility to choose among them possible targets, having similar profile for their contextual factors, similar input mixes and similar output values but being efficient. The current DMU #7 in black solid line is put at one for each input/output component, so the differences are in percentages.

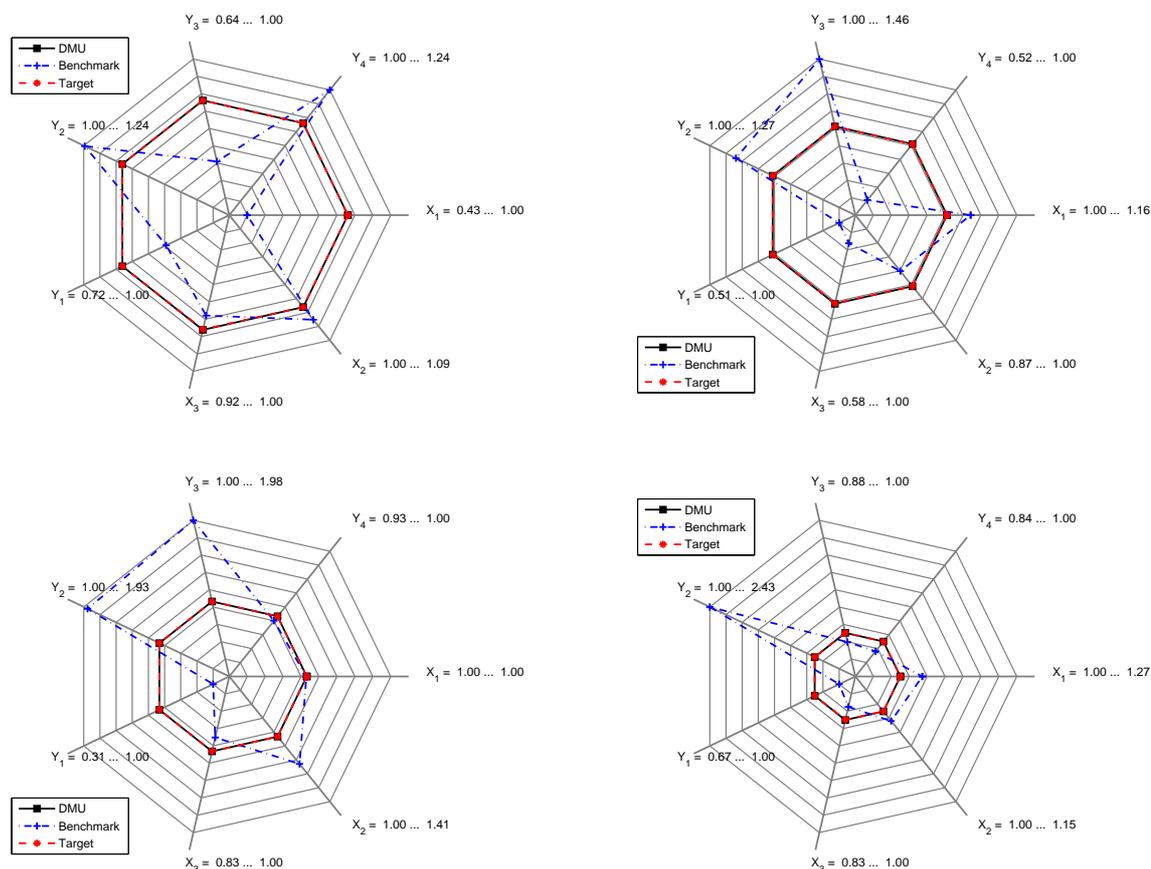


Figure 8: Radar Plots for unit #7. From left to right and top to bottom we have the benchmark efficient units #174, #46, #245 and #219.

6 Conclusions

In this paper we have introduced an empirical approach to endogenously determine the direction along which to assess the performance of DMUs without needing information about prices and without assuming any profit maximizing or cost minimizing behavior of the DMUs. When there are no economic or managerial reasons for imposing a specific path towards the efficient frontier, our approach is able to provide a data driven (local) direction which is able to account for the heterogeneity of DMUs and their contextual factors. The data

driven (local) direction provided by our approach lies between an egalitarian (global average) direction and an individual (specific) direction for each DMU and is able to detect irrelevant contextual factors. From a technical viewpoint, by introducing state of the art techniques in nonparametric regression for angular responses in an efficient frontier analysis framework, we detailed a procedure that allows us to automatically identify benchmarking peers for each DMU according to their specific context, without having to carry out an ad hoc clustering exercise. The usefulness of our approach has been illustrated with simulated as well as with real data sets on education and banking sectors.

A Appendix: Multidimensional Polar Coordinates

We present here precise and flexible formulae to transform cartesian multidimensional coordinates in polar coordinates and the inverse transformation. These results, in such general form are not so easy to find in the literature and this fix also the transformation we used in the paper.

A.1 Basic Definitions

There are many ways to define multidimensional polar coordinates. A popular one can be found in Anderson (2003, p.302). The definition is very general and is able to handle all the cases (with cartesian variables able to be in any orthant with some potential zero components). Let $x \in \mathbb{R}^p$, the modulus, or length, of x is given by $r = \|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_p^2}$. The polar coordinates are then completed by introducing the $p - 1$ components of the angle θ .

$$\begin{aligned}
 x_1 &= r \sin \theta_1 \\
 x_2 &= r \cos \theta_1 \sin \theta_2 \\
 x_3 &= r \cos \theta_1 \cos \theta_2 \sin \theta_3 \\
 &\vdots \\
 x_{p-1} &= r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{p-2} \sin \theta_{p-1} \\
 x_p &= r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{p-2} \cos \theta_{p-1},
 \end{aligned} \tag{A.1}$$

where $-\pi/2 < \theta_j \leq \pi/2$ for $j = 1, \dots, p - 2$ and $-\pi < \theta_{p-1} \leq \pi$. Note that the particular case $p = 2$ is covered by the above definition; it particularizes as

$$\begin{cases} x_1 = r \sin \theta \\ x_2 = r \cos \theta, \end{cases} \tag{A.2}$$

with $-\pi < \theta \leq \pi$ for covering all the 4 quadrants of $x \in \mathbb{R}^2$ (see Figure 9).

The above formulae can be seen as allowing to recover the cartesian variables $x \in \mathbb{R}^p$ from the knowledge of the polar coordinates (r, θ) where θ has $p - 1$ components with range defined the line after (A.1). We denote through the paper this transformation by $x = \phi(r, \theta)$.

A.2 The inverse transform

Let us start with the 2-dimensional problem. It can be seen from (A.2) that the angle θ has to satisfy the relation

$$\theta = \arctan(x_1/x_2), \tag{A.3}$$

but we have to cover the 4 quadrants depending on the values of (x_1, x_2) . The inverse function ‘arctan’ has to be defined accordingly.⁵

When $p > 2$, we first note from (A.1) that for all $j = 1, \dots, p - 2$,

$$\begin{aligned} \|x\|_{(j)} &:= \sqrt{x_{j+1}^2 + x_{j+2}^2 + \dots + x_p^2} = |r| |\cos \theta_1| |\cos \theta_2| \dots |\cos \theta_{p-2}| \\ &= r \cos \theta_1 \cos \theta_2 \dots \cos \theta_{p-2}, \end{aligned} \quad (\text{A.4})$$

where the last equality holds because $r > 0$ and $\cos \theta_j > 0$ since $-\pi/2 < \theta_j \leq \pi/2$ for $j = 1, \dots, p - 2$. Now we obtain directly the values of the first $p - 2$ angles, $j = 1, \dots, p - 2$:

$$\theta_j = \arctan \left(\frac{x_j}{\|x\|_{(j)}} \right), \quad (\text{A.5})$$

where the inverse function arctan is defined so that it takes its values in the range $[-\pi/2, \pi/2]$ (see the footnote 5 for practical use with Matlab or R). The last angle is simply given by

$$\theta_{p-1} = \arctan(x_{p-1}/x_p), \quad (\text{A.6})$$

where here, we use the version of arctan having range $[-\pi, \pi]$, as required by (A.1).

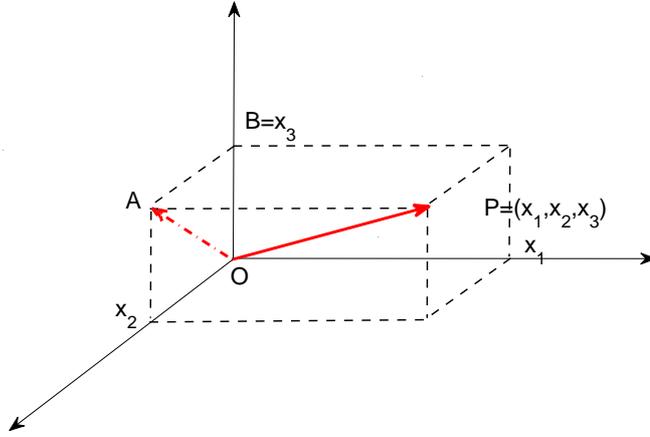


Figure 9: *An illustration of polar coordinates in 3-D: The point $P = (x_1, x_2, x_3)$ with modulus $r = |OP|$. The angles are $\theta_1 = \widehat{POA}$ and $\theta_2 = \widehat{BOA} = \arctan(x_2/x_3)$, we have $|AP| = x_1 = r \sin(\theta_1)$ and $|OA| = \sqrt{x_2^2 + x_3^2} = r \cos(\theta_1)$.*

⁵The Matlab (and R) build-in function `atan(a)` will give an angle η in $[-\pi/2, \pi/2]$ with $a = \tan(\eta)$. To cover the range $[-\pi, \pi]$ we have to use the Matlab build-in function `atan2(x1, x2)` to obtain our angle θ as defined by (A.3). It must be noticed that indeed in our general p -dimensional formulation, x_1 corresponds to the sine of the angle and x_2 to the cosine whereas it is often the contrary for the simple 2-dimensional case.

B Appendix: Nonparametric Regression for Angular Responses.

Circular or *directional* or *angular* data in a plane or in the space have been analyzed in the statistical literature. When trying to regress angular data on real-valued predictors or regressors, classical methods are not appropriate. This is because averaging over the values of a variable is not suited when using angles. Therefore, we have to define the regression of the angle, say θ on a real exogenous regressor, say W . Parametric approaches include Gould (1969), Fisher and Lee (1992) and Presnell et al. (1998). In the latter, maximum likelihood techniques are involved, and so von Mises-Fisher distributions for multivariate angle are used to model random angles. These approaches involves very restrictive and very strong assumptions that are made to make the procedures easy to use. Much more flexible approaches have not been developed so much, until recently. Di Marzio et al. (2013) (and the references therein) provide a nice framework for our goal and all the required theoretical properties of the resulting estimators have been derived within it. We will adapt hence their approach in our context.

To summarize, we are in the setting where a random angle θ may depend on a set of real-valued variables $W \in \mathbb{R}$. In what follows, we measure angles in the range $[-\pi, \pi]$ and use the standard anti-clockwise direction as positive. We define the function m on \mathbb{R} with values in $[-\pi, \pi]$, as the minimizer of the expectation of the usual angular distance $d(\theta, m(W)) = 1 - \cos(\theta - m(W))$. Following Di Marzio et al. (2013), for any $w \in \mathbb{R}$ we define

$$m_1(w) = \mathbb{E}[\sin(\theta)|W = w] \tag{B.1}$$

$$m_2(w) = \mathbb{E}[\cos(\theta)|W = w], \tag{B.2}$$

and $g_j(w) = m_j(w)f_W(w)$, $j = 1, 2$. The minimizer of the risk $\mathbb{E}[d(\theta, m(W))]$ is then given by

$$m(w) = \arctan(g_1(w)/g_2(w)), \tag{B.3}$$

where the inverse function \arctan has domain $[-\pi, \pi]$ (see Footnote 5 for practical use with Matlab or R). This approach corresponds to the regression model

$$\theta = m(W) + \varepsilon, \tag{B.4}$$

where the random angle ε has zero mean, finite dispersion and is independent of W .

Now, if we have a sample of observations (W_i, θ_i) , $i = 1, \dots, n$, we can use regular nonparametric regression techniques (like local constant or local linear, see Li and Racine,

2007 or Pagan and Ullah, 1999 for details) for estimating the functions $m_j(w)$ and the density $f_W(w)$. We end up with the Di Marzo et al. (2013) nonparametric estimator

$$\widehat{m}(w) = \arctan(\widehat{g}_1(w), \widehat{g}_2(w)), \quad (\text{B.5})$$

where the two components \widehat{g}_j are given, e.g., by the smoother⁶

$$\widehat{g}_1(w) = \frac{1}{nh} \sum_{i=1}^n \sin(\theta_i) K\left(\frac{W_i - w}{h}\right) \quad (\text{B.6})$$

$$\widehat{g}_2(w) = \frac{1}{nh} \sum_{i=1}^n \cos(\theta_i) K\left(\frac{W_i - w}{h}\right), \quad (\text{B.7})$$

where $K(\cdot)$ and h are respectively the kernel and the bandwidth appearing in most of the nonparametric kernel smoothers. Kernels are symmetric densities with support on $[-1, 1]$ (like Epanechnikov or quartic kernels) or any other symmetric density with finite second moment (like the standard normal) and the bandwidth, that is the smoothing parameter, controls for the localization, tuning the bias and the variance of the estimators.

The statistical properties of the estimator are established in Di Marzio et al. (2013) for Kernel with support $[-1, 1]$. To summarize, they show that the asymptotic properties of \widehat{g}_j and of \widehat{m} are very similar to those shared by the classical nonparametric regression estimators. In particular, and as usual, these properties assume that $h \rightarrow 0$ with $nh \rightarrow \infty$ when $n \rightarrow \infty$. The selection of the optimal bandwidth that balance (bias)² and variance produces an asymptotic mean squared error (AMSE) of the order $O(n^{-4/5})$ which is standard in nonparametric regression. For well-behaved parametric approaches we generally achieve the better order $O(n^{-1})$, but at a cost of restrictive parametric assumptions.

In practice, the optimal bandwidth is selected by leave-one-out cross-validation, like in the classical cases, except that here we have to use the appropriate risk function. The bandwidth is determined by the value h which minimize the criterion

$$CV(h) = -\frac{1}{n} \sum_{i=1}^n [\cos(\theta_i - \widehat{m}^{(i)}(W_i))], \quad (\text{B.8})$$

where $\widehat{m}^{(i)}(W_i)$ is the estimator $\widehat{m}(\cdot)$ evaluated at W_i by using all the observations except the pair (W_i, θ_i) . We note that at a constant, CV is an empirical estimates of the risk.

When W is multidimensional, the technique can be adapted by using multivariate kernels. As suggested by Li and Racine (2007) we will use product kernels with a vector of

⁶We describe the estimator for the local constant case, because in our setup, we must ensure the resulting estimate of the angle is in the same quadrant as the original observations. Local linear fit, also described in Di Marzio et al. (2013), could provide in finite sample inappropriate quadrants.

bandwidths: this allows to handle automatically relevant and irrelevant regressors in a flexible way. Practically, when h is as big as the range of W , this means that W is irrelevant for the choice of the direction (see Hall, Li and Racine, 2007 for details).

To fix the notations, suppose $W \in \mathbb{R}^d$ and denote by (h_1, \dots, h_d) a vector of bandwidths. The two components \hat{g}_j are now given for any $w \in \mathbb{R}^d$ by

$$\hat{g}_1(w) = \frac{1}{n \prod_{j=1}^d h_j} \sum_{i=1}^n \sin(\theta_i) \prod_{j=1}^d K\left(\frac{W_i^j - w^j}{h_j}\right) \quad (\text{B.9})$$

$$\hat{g}_2(w) = \frac{1}{n \prod_{j=1}^d h_j} \sum_{i=1}^n \cos(\theta_i) \prod_{j=1}^d K\left(\frac{W_i^j - w^j}{h_j}\right), \quad (\text{B.10})$$

where for any vector $a \in \mathbb{R}^d$, a^j denotes its j th component. Then (B.5) gives the estimate $\hat{m}(w)$. The properties of the resulting estimator are similar to the ones described in Di Marzio et al. (2013), except that here the AMSE will be of the order $O(n^{-4/(d+4)})$ showing that the estimator suffers from the usual ‘‘curse of dimensionality’’ common in most nonparametric approaches. That is, when the dimension d increases, we need more data to reach the same precision.

When we have to regress several angles $(\theta^1, \dots, \theta^p)$ on a vector $W \in \mathbb{R}^d$, as will be the case when using polar coordinates for the input $x \in \mathbb{R}^p$, it is common to use the approach just described component by component, by regressing each θ_j on W , for $j = 1, \dots, p$. The same procedure applies, *mutatis mutandis*, when we work in the output space $y \in \mathbb{R}^q$.

References

- [1] Anderson T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, 3rd Edition, John Wiley & Sons, Inc., New Jersey.
- [2] Banker, R.D. and R.C. Morey (1986), Efficiency analysis for exogenously fixed inputs and outputs, *Operations Research*, 34(4), 513–521.
- [3] Bogetoft P. (1997), DEA-based yardstick competition: The optimality of best practice regulation, *Annals of Operations Research*, 73, 277–298.
- [4] Chambers, R. G., Chung, Y.H. and R. Fare (1996), Benefit and Distance Functions, *Journal of Economic Theory*, 70, 407–419.
- [5] Chambers, R.G., Chung, Y.H. and R. Färe (1998), Profit, Directional Distance Functions and Nerlovian Efficiency, *Journal of Optimization Theory and Applications*, 98, 351–364.

- [6] Charnes, A., Cooper W.W. and E. Rhodes (1978), Measuring the inefficiency of decision making units, *European Journal of Operational Research*, 2 (6), 429–444.
- [7] Charnes, A., Cooper, W.W. and Rhodes, E., 1981, Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through, *Management Science* 27, 668–697.
- [8] Daraio C., Simar L. (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and applications*, Springer, New York.
- [9] Daraio C., Simar L. (2014), Directional distances and their robust versions: Computational and testing issues, *European Journal of Operational Research*, 237, 358-369.
- [10] Debreu, G. (1951), The coefficient of resource utilization, *Econometrica*, 19:3, 273–292.
- [11] Deprins, D., Simar, L. and H. Tulkens (1984), Measuring labor inefficiency in post offices. In *The Performance of Public Enterprises: Concepts and measurements*. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland, 243–267.
- [12] Di Marzio, M., Panzera, A. and C.C. Taylor (2013), Non-parametric Regression for Circular Responses, *Scandinavian Journal of Statistics*, 40, 2, 238–255.
- [13] Dyson, R.G. (2004), Strategic development and SWOT analysis at the University of Warwick, *European Journal of Operational Research*, 152(3), 631640.
- [14] Emrouznejad, A., Parker, B. and G. Tavares (2008), Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA, *Socio Economic Planning Science*, 42, 151-157.
- [15] Farrell, M.J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society*, A(120), 253–281.
- [16] Färe, R., and S. Grosskopf (2000), Theory and Application of Directional Distance Functions, *Journal of Productivity Analysis*, 13, 93–103.
- [17] Färe, R., and S. Grosskopf (2013), Directional output distance function: endogenous directions based on exogenous normalization constraints, *Journal of Productivity Analysis*, 13, 93–103.
- [18] Färe, R., Grosskopf, S. and D. Margaritis, (2008), Efficiency and Productivity: Malmquist and more, in: *The measurement of productive efficiency and productivity*

- growth*, edited by Fried H.O., Lovell C.A.K. and Schmidt S.S., Oxford University Press, pp. 522–622.
- [19] Färe, R., Grosskopf, S. and G. Whittaker (2013), Directional output distance functions: endogenous directions based on exogenous normalization constraints, *Journal of Productivity Analysis*, vol 40, 31, 267–269.
- [20] Fisher, N.I. and A.J. Lee (1992), Regression Models for Angular Response, *Biometrics*, 48, 665–667.
- [21] Gould, A.L. (1969), A regression technique for angular variates, *Biometrics*, 25, 683–700.
- [22] Hall, P., Li, Q. and J.S. Racine (2007), Nonparametric estimation of regression functions in the presence of irrelevant regressors, *The Review of Economics and Statistics*, 89, 4, 784–789.
- [23] Kerstens K. and Van de Woestyne I. (2011), Negative data in DEA: a simple proportional distance function approach, *Journal of the Operational Research Society*, 62, 1413-1419.
- [24] Kerstens, K., Mounir, A. and I. Van de Woestyne (2012), Benchmarking mean-variance portfolios using a shortage function: the choice of direction vector affects rankings!, *Journal of the Operational Research Society*, 63, 1199-1212.
- [25] Kneip, A., Simar, L. and P.W. Wilson (2013), Testing Hypothesis in Nonparametric Models of Production. Discussion paper 2013/48, Institut de Statistique, UCL.
- [26] Koopmans, T.C.(1951), An analysis of production as an efficient combination of activities, in Koopmans, T.C. (ed) *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics, Monograph 13, John-Wiley, New-York.
- [27] Korhonen P., Tainio R., Wallenius J. (2001), Value efficiency analysis of academic research, *European Journal of Operational Research*, 130 (1), 121-132.
- [28] Li, Q. and J.S. Racine (2007), *Nonparametric Econometrics*, Princeton University Press, Princeton and Cambridge.
- [29] Pagan, A. and A. Ullah (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge.

- [30] Peyrache A., C. Daraio (2012), Empirical tools to assess the sensitivity of directional distance functions to direction selection, *Applied Economics*, 44, 933–943.
- [31] Presnell, B., Morrison, S.P. and R.C. Littell (1998), Projected Multivariate Linear Models for Directional Data, *Journal of the American Statistical Association*, 93, 443, 1068–1077.
- [32] Racine, J. S. (2008), Nonparametric Econometrics: A Primer, *Foundations and Trends[®] in Econometrics*, 3 (1), 1–88.
- [33] Shephard, R.W. (1970), *Theory of Cost and Production Function*. Princeton University Press, Princeton, New-Jersey.
- [34] Simar, L. (2003), Detecting Outliers in Frontiers Models: a Simple Approach, *Journal of Productivity Analysis*, 20, 391–424.
- [35] Simar, L. and A. Vanhems (2012), Probabilistic Characterization of Directional Distances and their Robust versions, *Journal of Econometrics*, 166, 342–354.
- [36] Simar L., Vanhems A., Wilson P.W. (2012), Statistical Inference for DEA Estimators of Directional Distances, *European Journal of Operational Research*, 220(3), 853–864.
- [37] Simar, L. and P.W. Wilson (2011), Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis*, 36, 33–53.
- [38] Simar, L. and P.W. Wilson (2014), Statistical Approaches for Nonparametric Frontier Models: A Guided Tour, in press, *International Statistical Review*.
- [39] Thanassoulis, E. Portela, M. C. S., Despic, O. (2008), Data Envelopment Analysis: The mathematical programming approach to efficiency analysis, in *The measurement of productive efficiency and productivity growth*, edited by Fried H.O., Lovell C.A.K. and Schmidt S.S., Oxford University Press, pp. 251–420.
- [40] Wilson, P. W. (1993), Detecting outliers in deterministic nonparametric frontier models with multiple outputs, *Journal of Business and Economic Statistics*, 11, 319–323.
- [41] Zofio J.L., Pastor J.T. and J. Aparicio (2013), The directional profit efficiency measure: on why profit inefficiency is either technical or allocative, *Journal of Productivity Analysis*, 40, 257–266.