INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

WORKING PAPER SERIES

Approximations and Inference for Nonparametric Production Frontiers

Cinzia Daraio [a]
Léopold Simar [b]

[a] Department of Computer, Control and Management Engineering A. Ruberti (DIAG), Sapienza University of Rome, Italy.
[b] Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA), LIDAM, Université Catholique de Louvain, Belgium.

2022/14 May 2022
ISSN(ONLINE) 2284-0400

Approximations and Inference for Nonparametric Production Frontiers

Cinzia Daraio*
daraio@diag.uniroma1.it

Léopold Simar§
leopold.simar@uclouvain.be

May, 2022

Abstract

Nonparametric methods have been widely used for assessing the performance of organizations in the private and public sector. The most popular ones are based on envelopment estimators, like the FDH or DEA estimators, that estimate the attainable sets and its efficient boundary by enveloping the cloud of observed units in the appropriate input-output space. The statistical properties of these flexible estimators have been established. However these nonparametric techniques do not allow to make sensitivity analysis of the production outputs to some particular inputs, or to infer about marginal products and other coefficients of economic interest. On the contrary, parametric models for production frontiers allow richer and easier economic interpretation but at a cost of restrictive assumptions on the data generating process. In addition, the latter rely mostly on regression methods fitting the center of the cloud of observed points. In this paper we offer a way to avoid these drawbacks and provide approximations of these coefficients of economic interest by "smoothing" the popular nonparametric estimators of the frontiers. Our approach allows to handle fully multivariate cases. We describe the statistical properties for both the full and the partial (robust) frontiers. We consider parametric but also flexible approximations based on local linear tools providing local estimates of all the desired partial derivatives and we show how to deal with environmental factors. An illustration on real data from European Higher Education Institutions (HEI) shows the usefulness of the proposed approach.

Key Words: Nonparametric production frontiers, DEA, FDH, partial frontiers, directional distances, linear approximations, local linear approximations.

JEL Classification: C1,C14, C13;

1 Introduction and Basic Notations

The theory of production and efficiency analysis examine how firms or production units transform their inputs (like, labour, energy and capital) into outputs, that is the quantities of goods or services that are produced by the units. These analyses have been developed for manufacturing plants, for business, profit oriented firms but also for schools, non-profit organizations, providers of public goods and services to examine how these units organize their production process.

For achieving this goal, production economics has developed a rich theory since the pioneering work of Koopmans (1951) and Debreu (1951). Farrell (1957) and Afriat (1972) were among the first to investigate these issues from an empirical perspective. The efficient production frontier is defined in the appropriate input-output space as the locus of the optimal combinations of the inputs and the outputs. Formally, let the attainable set, i.e. the set of technically feasible combinations of inputs and outputs, be defined as

$$\Psi = \{(x, y) \in \mathbb{R}^{p+q} | \ x \text{ can produce } y\}. \tag{1.1}$$

This set shares the usual characteristics coming from economic theory (see e.g. Shephard, 1970), the efficient boundary (frontier) of this set is the set of efficient combinations of inputs and outputs

$$\Psi^{\partial} = \{(x, y) \in \mathbb{R}^{p+q} | (\gamma^{-1}x, \gamma y) \notin \Psi, \forall \gamma > 1\}. \tag{1.2}$$

*Department of Computer, Control and Management Engineering A. Ruberti (DIAG), Sapienza University of Rome, via Ariosto 25, 00185, Rome, Italy. The financial support of Sapienza project awards n. RM11916B8853C925 and n. RM12117A8A5DBD18 is gratefully acknowledged.

§Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA), LIDAM, Université Catholique de Louvain, Voie du Roman Pays 20, B 1348 Louvain-la-Neuve, Belgium.

There are several ways for measuring the efficiency of a production plan (x, y) as the distance from this boundary. In this paper we mainly focus on the very flexible directional distances measures (see Chambers et al., 1998), defined as

$$\delta(x,y) = \sup\{\delta | \, (x - \delta d_x, y + \delta d_y) \in \Psi\}, \tag{1.3}$$

where $d_x \in \mathbb{R}_+^p$ and $d_y = \mathbb{R}_+^q$. So the distance is measured along a path determined by a direction vector $d' = (-d_x', d_y')$ in an additive way. Clearly if $(x, y) \in \Psi$, $\delta(x, y) \geq 0$ and if (x, y) lies on the efficient frontier (1.2), $\delta(x, y) = 0$. The Farrell-Debreu oriented radial distances and the radial hyperbolic distances (Färe et al., 1985) can be recovered as special cases (see below in Section 3.5). It will be useful below to denote as $w^\partial = (x^\partial, y^\partial)$, the projection of $w = (x, y)$ on the efficient frontier in the direction d, i.e. $w^\partial = w + \delta(w)d$. Component by component

$$x^\partial = x - \delta(x,y)d_x, \text{ and } y^\partial = y + \delta(x,y)d_y. \tag{1.4}$$

Note that distance functions satisfy the "translation" property:

$$\delta(w + \eta d) = \delta(w) - \eta, \text{ for all } \eta \in \mathbb{R}. \tag{1.5}$$

Similarly, partial frontiers (as initiated by Cazals et al, 2002, Aragon et al., 2005, Daouia and Simar, 2007) can also be considered providing order-m frontiers and order-α quantile frontiers. This allows to define $\delta_m(x, y)$ and $\delta_\alpha(x, y)$ which by construction are smaller than $\delta(x, y)$, unless $m \to \infty$ and $\alpha \to 1$ (see Simar and Vanhems, 2012). These efficiency measures benchmark a production plan (x, y) against less extreme frontiers, and so share robustness properties, robustness against outliers or extreme observations. For instance for the order-m case, the "partial frontier" points are defined as

$$x_m^\partial = x - \delta_m(x,y)d_x, \text{ and } y_m^\partial = y + \delta_m(x,y)d_y, \tag{1.6}$$

and similarly for the order-α frontiers.

In practice, the objects defined above are unknown and must be estimated from a random sample of observations $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$. The paradigm considered here is the so called "deterministic frontiers approach" where the Data Generating Process (DGP) specifies that

$$\text{Prob}\big((X,Y) \in \Psi\big) = 1, \tag{1.7}$$

so that the support of the joint density $f(x, y)$ of (X, Y) is a subset of Ψ.[1] The popular nonparametric estimators of Ψ are based on envelopment estimators, like Free Disposal Hull (FDH) or Data Envelopment Analysis (DEA) estimators. From these estimators, it is easy to define for any (x, y), $\widehat{\delta}_n(x, y)$, a nonparametric estimators of $\delta(x, y)$. Estimators of partial efficiency scores $\widehat{\delta}_{m,n}(x, y)$ and $\widehat{\delta}_{\alpha,n}(x, y)$ are derived in Simar and Vanhems (2012) and share very nice and attractive properties (see below). Practical ways to compute these estimators are described in Daraio et al. (2020).[2]

These nonparametric estimators are widely used because they rely on very few assumptions: no particular shape for the attainable set and its frontier (only free disposability for the FDH and in addition, convexity of the attainable set for the DEA) and no particular distributional assumption for the distribution of (x, y) on Ψ.[3] The statistical properties of these envelopment estimators have been established and inference is available (see Simar and Wilson, 2013 and 2015, for recent surveys). Their drawback is that the results are difficult to interpret in terms of the sensitivity of the production of certain output to particular inputs, marginal rates of substitution between inputs, marginal rates of transformation between outputs and so on.

On the other hand, parametric models have been proposed. Most of the parametric models are based on standard regression methods taking into account for the one sided error term: shifted OLS, corrected OLS and MLE (see e.g. Kumbhakar and Lovell, 2000). These parametric models offer an easier and richer economic interpretation of the production process but at a risk of misspecification resulting in a risk of inconsistent estimates. In particular most of these techniques imply the specification of a parametric family for the density of the one-sided stochastic error term. In addition, as noted in Florens and Simar (2005) these parametric regression methods fit the center of the cloud of observed points rather than the shape of the cloud near its efficient boundary.

Hence, Florens and Simar (2005) have suggested to approximate nonparametric estimators of the frontier by parametric models, to allow for richer economic interpretation of the resulting model and reach a better fit

[1]This is opposed to the "stochastic frontier models", often denoted by the acronym SFA (Stochastic Frontier Analysis), which allows for noise in the DGP, and so the support of $f(x, y)$ may extend outside Ψ.

[2]Daraio et al. (2020) provide the Matlab codes for the needed computations. These are also available in the FEAR software package in R, introduced by Wilson (2008).

[3]Free disposability means that for any $\tilde{x} \geq x$ and any $\tilde{y} \leq y$, if $(x, y) \in \Psi$ then $(\tilde{x}, \tilde{y}) \in \Psi$.

of the efficient boundary. They show how a two-stage method provides an estimation of the parametric model that overcomes most of the drawbacks of traditional one-stage regression models, in particular avoiding to specify any parametric family of densities for the stochastic part of the model and fitting a cloud of points near the efficient boundary. They illustrate the advantage of their method compared with the traditional methods by various illuminating simulated examples.

As most of the parametric approaches, Florens and Simar (2005) is restricted to the case of univariate response (a cost function for univariate x or production function for univariate y). Their analysis is done for full frontier where FDH estimators are used in the first stage (here only consistency of the resulting estimators is proven) and for order-m frontiers (where asymptotic Gaussian inference is available). Daouia et al. (2008) extend the latter to order-α frontiers.

For fully multivariate setup, parametric models for radial efficiency measures, have been proposed. These involve regressing one of the outputs on the inputs and the other outputs divided by the dependent variable. Hence, this leads to endogeneity problems that are either ignored or solved by adding restrictive assumptions (see e.g. Grosskopf et al., 1997, Coelli and Perelman, 1999). We propose in this paper a way to overcome these drawbacks. The difficulty of the multivariate setup, where $r = p + q$, is that the efficient frontier in the (x, y)-space is an $(r - 1)$-manifold. One way to overcome this difficulty is to express, in an appropriate coordinates system, the efficient frontier as a scalar-valued, $(r - 1)$ variate function. In a nutshell, the transformation is a rotation of the (x, y) coordinate system so that one coordinate is parallel to the chosen direction vector d, and the others are orthogonal to d.

This transformation has been used to provide natural ways to build estimators of SFA in a fully multivariate setup in Simar and Wilson (2021) (hereafter SW). SFA models have their advantages (allowing noisy data) but also their drawbacks (identification issues, possibility of wrong skewness of the composite error term, see SW for a discussion and suggested solutions). In our paper, we will see how to adapt the DGP described in SW to our setup of deterministic frontier models, we are also able to derive new explicit relations between the characteristics in both coordinate systems. Then we will show how the strategy of Florens and Simar (2005) and Daouia et al. (2008) can easily be adapted. This allows us to capture the shape of the cloud of points near its efficient boundary, without specifying any parametric family of densities for the stochastic part of the model and this avoids the endogeneity problems mentioned above. The idea is to "smooth" by some appropriate models the usual nonparametric estimators of the frontier. We present our approach in a directional distance setup, but we show how to adapt the approach to any other measure of efficiency (hyperbolic, input or output radial oriented). We present first a parametric approximation (in the lines of Florens and Simar, 2005). Then we show that more flexible local linear approximations are also easy to handle, providing approximations of all the coefficients of economic interest, including derivatives, despite its nonparametric nature. Finally, we illustrate how the approach can easily be extended to deal with environmental factors.

The paper is organized as follows. In the next section, we describe the main objective of our paper and its contribution to the existing literature. Section 3 illustrates the underlying statistical model (i.e. the DGP) and the transformation necessary to implement the fully multi-inputs multi outputs case. Section 4 introduces the best approximation and its estimation. Section 5 details linear approximation models (including the special cases of Cobb-Douglas and Translog) and the related statistical properties. Section 6 presents the local linear approximation and the related inference. Section 7 explains how we can extend the approach when environmental factors may influence the frontier. Section 8 reports numerical illustrations on simulated and real data and Section 9 concludes the paper.

2 Aim and Contribution

In this paper we address a long-standing issue in efficiency analysis, which is whether to choose a parametric or nonparametric approach to perform an efficiency analysis. As we have seen in the previous section, both approaches (parametric and nonparametric) have merits (pros) and demerits (cons) and do not always offer satisfactory solutions for users of efficiency techniques interested in the results, their robustness, and their interpretability in the application context. Our proposal is not to leave the users of efficiency techniques with the Hamletic choice of "Parametric or non-parametric, this is the problem..." but to provide an approach that combines the two approaches (parametric and non-parametric) seeking a balance between the two, as an old Latin saying goes *in medio stat virtus* (meaning "virtue lies in the middle").

We propose a two-stage approach that combines the flexibility of a first-stage based on nonparametric models (like FDH, DEA and partial efficiency measures) with a flexible second-stage based on a parametric adjustment of the nonparametric efficient frontier estimated in the first-stage. Compared to existing approaches, our contribution adds (i) a parametric adjustment of the efficient frontier that represents a "smoothing" of the nonparametric frontier estimated in the first-stage that does not assume any arbitrary choice of the frontier function; (ii) a smoothing parametric adjustment that captures the shape of the nonparametrically estimated efficient frontier instead of capturing the center of the cloud of points as is usually done in existing approaches; (iii) an approach in which it is not necessary to specify a functional form of the inefficiency (such as a half normal, exponential or gamma); (iv) a treatment of fully multi-output and multi-input cases; (v) a flexible underlying efficiency model based on directional distances where one can choose the desired benchmark direction; (vi) guidelines for statistical inference for both the full and the partial (robust) frontiers approximations, and (vii) a flexible approximation based on local linear models.

To show the usefulness of our approach we report an illustration in the field of higher education where the assumptions underlying parametric models have longed been challenged.

According to Hanushek (1979) who has questioned parametric models in education, the measurement of educational performance and its determinants is affected by a lack of conceptual clarity and severe analytical problems, including the consideration of multiple output in isolation without taking into account the interactions among them in the production. The estimation raises a series of conceptual and statistical problems related to the choice of the functional form, the level of aggregation, selection effects and causation and multicollinearity to cite a few. In standard microeconomics, a production function describes the maximum achievable (feasible) output for given inputs. Differently with respect to inputs-outputs characterization, production functions defined as above, implicitly assume a profit maximizing behavior of the units.[4] Later, Figlio (1999), Dewey et al. (2000) and Baker (2001), put further the inquiry about the results obtained by traditional education production functions based on parametric approaches and showed that without imposing restrictive assumptions different results were obtained.

Stock (2010) traces the development of econometric models from the traditional ones of the eighties, mostly parametric and characterized by a linear functional form, to more recently developed nonparametric ones, thanks to the development of the computer power and the advancements of mathematical and statistical research. He identifies one of the causes of the development of nonparametric models in dissatisfaction towards traditional parametric models that were not always a good approximation.[5]

In this paper, we take the next step of what Stock (2010) described, and propose to combine the new nonparametric methods (i.e. robust directional distances) and related inferential approaches (i.e., local linear models) to obtain more robust coefficients and quantities of economic interest that do not rely on arbitrary choices of functional forms.

3 The Statistical Model and the Transformation

3.1 The DGP

The statistical model describes the way the observations are obtained, i.e. the DGP. We adapt the SFA model proposed in SW (Assumptions 2.2 and 2.3 in SW) to our setup. We assume that the production process generates efficient but unobserved production plans. Then we describe how the observed production plans are generated due to inefficiency. Formally, the DGP generates random optimal production plans on the effcient boundary Ψ^∂ via some probability mechanism, providing identically, independently distribted (iid) values $W_i^\partial = (X_i^\partial, Y_i^\partial)$, $i = 1, \ldots, n$.

We then assume that the random deviations from the efficient frontiers providing the observed production plans are along the direction vector d. In our approach, d is fixed and non-stochastic and the same for all organizations. The observed input-output pairs are denoted by $W_i = (X_i, Y_i)$ and defined by the model $W_i = W_i^\partial - \delta_i d$, i.e., component by component,

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{bmatrix} X_i^\partial \\ Y_i^\partial \end{bmatrix} - \delta_i \begin{bmatrix} -d_x \\ d_y \end{bmatrix}, \tag{3.1}$$

[4]Hanushek (1986) states: "Production functions are generally assumed to be known precisely by decision makers, to involve only a few inputs that are measured perfectly, and to be characterized by a deterministic relationship between inputs and outputs (that is, a given set of inputs always produces exactly the same amount of output). Furthermore, it is assumed that all inputs can be varied freely. The realities of education differ considerably from such pedagogical assumptions. Indeed, the production function is unknown (to both decision makers and researchers) and must be estimated using imperfect data; some important inputs cannot be changed by the decision maker; and any estimates of the production function will be subject to considerable uncertainty. Perhaps the largest difference between applying production functions to education and to other industries, however, has been in its immediate application to policy considerations. Statistical estimates of educational production functions have entered into a variety of judicial and legislative proceedings and have formed the basis for a number of intense policy debates".

[5]Stock (2010) states as follows: "The past three decades have seen significant changes in the tools of econometrics, many motivated by a desire to minimize the effect of whimsical assumptions on inference about the object of interest. By whimsical I mean arbitrary assumptions that are subsidiary to the empirical purpose at hand, but which affect inference about the causal effect of interest. The new tools provide reliable inference without implausible subsidiary assumptions (Stock, 2010, pg. 84-85)."

where the δ_i are conditionally to W_i^∂ independent with $\delta_i \mid W_i^\partial \sim D_+\big(\eta(W_i^\partial)\big)$ and $D_+(\cdot)$ being some one-sided distribution on $\mathbb{R}_+$ characterized by finite dimensional parameters $\eta(W_i^\partial)$. We will come back below to this distribution, we only assume for now that the corresponding density is strictly positive at zero (as in Park et al, 2000 for FDH and Kneip et al., 2008 for DEA) to guarantee the rates of convergence used below for the envelopment estimators.

These assumptions ensure we have a random sample of observations $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$ that we can use to derive the envelopment estimators like FDH, DEA and also their robust versions: order-m and order-α. The resulting estimator $\widehat{\delta}(x, y)$ shares the known properties described in the literature (mainly Simar and Vanhems, 2012, Simar et al., 2012). To summarize, for the full frontiers, we have (under mild regularity conditions) for a fixed point of interest (x, y), as $n \to \infty$,

$$n^\kappa\Big(\widehat{\delta}_\bullet(x,y) - \delta(x,y)\Big) \xrightarrow{\mathcal{L}} F_\bullet(\xi_{x,y}), \tag{3.2}$$

where κ determines the rate of convergence and depends on the assumptions on Ψ and the chosen estimator with "$\bullet$" representing FDH or DEA and $F_\bullet(\xi_{x,y})$ is a non-degenerate distribution depending on some unknown parameters. Under the free disposability assumption only, the FDH estimator has to be used and $\kappa = 1/(p + q)$, if we add the convexity assumption then we can also use the DEA estimator with $\kappa = 2/(p + q + 1)$. Typically the points of interest are the observations and so we may obtain estimators $\widehat{\delta}_i$. Kneip et al. (2015) derive central limit theorems for functions of $\widehat{\delta}_i$. The achieved rates of convergence given by n^κ illustrate the curse of dimensionality, common in nonparametric estimation: if $p + q$ increases, we loose precision in the estimation and we may be far below the usual $\sqrt{n}$ rate of convergence usually reached by parametric estimators.

The partial robust frontiers share two attractive properties: (i) by construction, they are less extreme than the full frontier and so their estimators will not envelop all the data hence they are more robust to extreme data points and outliers and (ii) they are asymptotically normally distributed with mean zero around the true values, with the parametric rate $\sqrt{n}$. We have, as $n \to \infty$,

$$\sqrt{n}\Big(\widehat{\delta}_\star(x,y) - \delta_\star(x,y)\Big) \xrightarrow{\mathcal{L}} N(0, \sigma^2_{\star,x,y}), \tag{3.3}$$

where "$\star$" represents the order-m (e.g. $\delta_m(x, y)$) or the order-α (e.g. $\delta_\alpha(x, y)$) case, and $\sigma^2_{\star,x,y}$ has a known expression depending on the chosen measure $\star$ and of the characteristics of the DGP. So inference with these estimators is much more easy. We will come back to all these nonparametric estimators below.

3.2 The rotation

A natural way to transform the coordinates $w = (x, y)$ in the original space into a new system where the distance to the frontier (defined in (3.1)) can be expressed by a scalar-valued equation, is to rotate the coordinates so that in the new system, one coordinate is parallel to d and the remaining $r - 1$ coordinates are orthogonal to d. As described in Simar and Wilson (2021) this is achieved by considering an arbitrary, but fixed, orthonormal basis for the direction vector d.[6] Let $S_d = [s_1 \ \dots \ s_{r-1}]$ be such $r \times (r - 1)$ matrix with $s_j's_j = 1$, $s_j's_k = 0$ for $j \neq k$ and $s_j'd = 0$ for $j = 1, \dots, r - 1$. Clearly $S_d'S_d = I_{r-1}$ and $S_d'd = 0_{r-1}$.[7] Now we can define the $r \times r$ rotation matrix

$$R_d = \begin{bmatrix} S_d' \\ d'/||d|| \end{bmatrix}, \text{ and its transpose } R_d' = [S_d \quad d/||d||], \tag{3.4}$$

where $|| \cdot ||$ denotes the L_2-norm. Clearly R_d is an orthogonal matrix, implying $R_d^{-1} = R_d'$. We note that $R_d d = [0_{r-1}' \quad ||d||]'$.

We consider now the linear transformation from $\mathbb{R}^r$ to $\mathbb{R}^r$ given by

$$g_d : w \mapsto t = R_d w, \tag{3.5}$$

which can easily be inverted, i.e. $w = R_d' t$. To see the consequence of this transformation, we partition $t' = (v' \ \ u)$ where $v = S_d' w$ and $u = d'w/||d||$, the rotation puts the coordinate u in the direction d and the

[6] As noted by SW, an orthonormal basis for d, the matrix S_d, is not unique, but this does not create any problem provided S_d is fixed after it is selected. Most of the statistical packages have a build-in function to obtain S_d. In case, an easy to program algorithm is described in Jeong and Simar (2006).

[7] I_k denotes the identity matrix of order k and 0_k a k-dimensional column vector of zeros.

$r-1$ remaining coordinates v are orthogonal to d (and hence to the u-axis). In the new coordinate system, the attainable set Ψ is represented by

$$\Gamma_d = \{t \in \mathbb{R}^r \mid t = g_d(w), w \in \Psi\}. \tag{3.6}$$

The efficient frontier can now be represented in terms of the scalar valued function

$$\phi(v) = \sup\{u \mid t = (v' \quad u)' \in \Gamma_d\}, \tag{3.7}$$

which permits to describe the attainable set in terms of this function

$$\Gamma_d = \{t = (v' \quad u)' \in \mathbb{R}^r \mid u \leq \phi(v)\}. \tag{3.8}$$

Figure 1, in Section 8.1.2 below, illustrates this rotation for the case $p = q = 1$.

Applying the rotation to the observations $W_i = (X_i, Y_i) \in \mathcal{X}_n$ yields the random sample of values $\{(V_i, U_i)\}_{i=1}^n$. To be explicit

$$\begin{bmatrix} V_i \\ U_i \end{bmatrix} = R_d W_i = \begin{bmatrix} S_d' W_i \\ ||d||^{-1} d' W_i \end{bmatrix}, \tag{3.9}$$

and the inverse relation between the observations is

$$W_i = R_d' \begin{bmatrix} V_i \\ U_i \end{bmatrix} = \left[S_d V_i + ||d||^{-1} d\, U_i\right]. \tag{3.10}$$

Hence our model (3.1) is transformed into

$$\begin{bmatrix} V_i \\ U_i \end{bmatrix} = \begin{bmatrix} V_i^\partial \\ U_i^\partial \end{bmatrix} - \delta_i \begin{bmatrix} 0_{r-1} \\ ||d|| \end{bmatrix},, \tag{3.11}$$

or component by component

$$\begin{array}{l} V_i = V_i^\partial \\ U_i = U_i^\partial - ||d||\delta_i \end{array}, \tag{3.12}$$

or simply due to (3.7)

$$U_i = \phi(V_i) - ||d||\delta_i, \tag{3.13}$$

where the heteroskedastic nature of δ in (3.1) can now be expressed in terms of V_i, since $W_i^\partial = R_d'[V_i' \quad \phi(V_i)]'$. So (3.13) provides the scalar-valued equation to be estimated.

3.3 Some relations between the two spaces

We will see in Section 4 that we can provide approximations for the function $\phi(v)$ and for the $(r-1)$-vector of partial derivatives, $\partial\phi(v)/\partial v'$ at any value v in its range (we assume smoothness of the frontier to ensure the existence of the partial derivatives below). So it is important to see if we can recover from these, the properties of the frontier and of the distance function in the original units.

3.3.1 Characteristics of the frontier

Since $t = (v' \quad u)' = R_d w$ and $w = R_d' t$, the frontier surface in the $w = (x, y)$-space is a $(r-1)$-manifold that can be obtained as

$$w^\partial = \begin{bmatrix} x^\partial \\ y^\partial \end{bmatrix} = [S_d \quad d/||d||] \begin{bmatrix} v \\ \phi(v) \end{bmatrix}, \tag{3.14}$$

i.e. a mapping from $\mathbb{R}^{r-1}$ to $\mathbb{R}^r$. One of the interest in this multivariate setup is to characterize the hyperplane tangent to this surface at some given point w_0^∂ and to derive various parameters of economic interest. This can be done as follows.

In the t-space, the hyperplane tangent at the frontier $\phi(v)$ at a frontier point $t_0^\partial = (v_0' \quad u_0^\partial)'$ with $u_0^\partial = \phi(v_0)$ is given by the equation

$$c_t'(t - t_0^\partial) = 0, \text{ where } c_t' = [\nabla'\phi(v_0) \quad -1], \tag{3.15}$$

with $\nabla\phi(v_0) = \left[\frac{\partial\phi(v)}{\partial v}\right]_{v=v_0}$ being the $(r-1)$-vector of the gradients of $\phi(v)$ evaluated at v_0. Now in the w-space this hyperplane has equation $c_t' R_d(w - w_0^\partial)$, i.e.

$$c_w'(w - w_0^\partial) = 0, \text{ where } c_w = R_d' \begin{bmatrix} \nabla\phi(v_0) \\ -1 \end{bmatrix} \text{ and } v_0 = S_d' w_0^\partial. \tag{3.16}$$

Clearly the vectors c_t and c_w are given at a multiplicative constant ($\neq 0$).[8]

Hence, the partial derivatives of the frontier surface in the (x, y)-space at the point w_0^∂ can be written as

$$\left.\frac{\partial w_\ell}{\partial w_k}\right|_{w=w_0^\partial} = -\frac{c_k}{c_\ell}, \tag{3.17}$$

provided $c_\ell \neq 0$. This can be used to derive the marginal products $\partial Y_\ell/\partial X_k$, marginal rates of substitution $\partial X_\ell/\partial X_k$ and the marginal rates of transformation $\partial Y_\ell/\partial Y_k$. If $c_\ell = 0$, then the derivative in (3.17) is not defined, however, this indicates that w_ℓ has no effect on the frontier at this particular frontier point. So, if $c_k \neq 0$

$$\left.\frac{\partial w_k}{\partial w_\ell}\right|_{w=w_0^\partial} = 0. \tag{3.18}$$

3.3.2 Distance function

We can also recover the distance function $\delta(x, y)$ in the original units and its partial derivatives. In the transformed space, we have from (3.13)

$$\delta(v, u) = ||d||^{-1}\big(\phi(v) - u\big), \tag{3.19}$$

so in the original $w = (x, y)$-space we have

$$\delta(w) = ||d||^{-1}\big(\phi(S_d' w) - d'w/||d||\big). \tag{3.20}$$

It is easy to check that this distance function satisfies the translation property (1.5). This confirms that the transformation has preserved all the desired properties of the original distance function. The reader can verify that we would obtain the same relation for $\delta(w)$ by writing $w^\partial = w + \delta(w)d$. Since from (3.14) we have $w^\partial = S_d v + d\phi(v)/||d||$ then by plugging $v = S_d' w$ we obtain (3.20).[9]

Note that we can also obtain the partial derivatives

$$\frac{\partial\delta(w)}{\partial w} = ||d||^{-1}\left(S_d \frac{\partial\phi(v)}{\partial v'} - \frac{d}{||d||}\right), \tag{3.21}$$

where $v = S_d' w$. Now we know that the frontier points w^∂ are characterized by the equation $\delta(w^\partial) = 0$. Hence the $(r-1)$-manifold describing the frontier in (3.14) can also be given by

$$\phi(S_d' w^\partial) - d' w^\partial/||d|| = 0, \tag{3.22}$$

which is nothing else than rewriting, in terms of w^∂, the equation for frontier points in the t-space, $\delta(v, u) = 0$, or equivalently $\phi(v) - u = 0$.

3.4 Two simple examples

Before going into simple examples where all the elements described above have simple analytic expressions, we first remark that in the particular case of univariate output ($q = 1$) and output orientation ($d = [0_p' \ \ 1]'$), the transformation is trivial, $R_d = I_r$, so there is no rotation: $v = x$ and $u = y$, and we are back to the univariate response case described in Florens and Simar (2005).

[8]Note that we provide by (3.16) an explicit expression for c_w. It is easy to check that c_w is a basis of the null space of the Jacobian of the transformation defined in (3.14), which was the way chosen by Simar and Wilson (2021) to characterize implicitly the vector c_w.

[9]To see this, we use the fact that R_d is orthogonal, so that $R_d' R_d = S_d S_d' + dd'/||d||^2 = I_r$.

3.4.1 A toy example

We consider the simplest one input and one output case ($p = q = 1$) with the direction vector $d = [-1 \;\; 1]'/\sqrt{2}$, so that $||d|| = 1$ to simplify the notations. Here we can choose $S_d = [1 \;\; 1]'/\sqrt{2}$ so that

$$R_d = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Here $\nabla\phi(v)$ is the (univariate) derivative of $\phi(v)$ and $c_t = [\nabla\phi(v) \;\; -1]'$, from this we derive

$$c_w = \frac{1}{\sqrt{2}} \begin{bmatrix} \nabla\phi(v) + 1 \\ \nabla\phi(v) - 1 \end{bmatrix} = \begin{bmatrix} c_x \\ c_y \end{bmatrix}.$$

Hence we have the relation between the derivative of the frontier in the original space and the characteristics in the t space given by the following by using (3.17) : at some frontier point w_0^∂,

$$\left.\frac{\partial y}{\partial x}\right|_{w_0^\partial} = -\frac{c_x}{c_y} = \frac{1 + \nabla\phi(v_0)}{1 - \nabla\phi(v_0)}, \tag{3.23}$$

where $v_0 = (x_0^\partial + y_0^\partial)/\sqrt{2} = (x_0 + y_0)/\sqrt{2}$.

Let us now particularize to an even simpler case where the frontier is linear and so, $y^\partial = a + bx^\partial$. In this case some simple algebra leads to the distance function $\delta(x, y) = (a + bx - y)\sqrt{2}/(1 + b)$. Applying the transformation R_d, it is easy to check that $v = (x + y)/\sqrt{2}$ and $u = (y - x)/\sqrt{2}$. For the frontier points we have also (remember that $v^\partial \equiv v$)

$$\begin{bmatrix} v \\ u^\partial \end{bmatrix} = R_d \begin{bmatrix} x^\partial \\ y^\partial \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} a + (b+1)x^\partial \\ a + (b-1)x^\partial \end{bmatrix}.$$

Then from this system of equations (the first equation gives x^∂ as a function of v that can be plugged in the second equation) we obtain the following explicit expression for $\phi(v)$

$$u^\partial = \phi(v) = \frac{\sqrt{2}a}{b+1} + \frac{b-1}{b+1}v,$$

leading to the derivative $\nabla\phi(v) = (b-1)/(b+1)$.

Now we let the reader verify that from the only knowledge of R_d, $\phi(v)$ and $\nabla\phi(v)$ we can recover all what we need in the (x, y)-space (see also the linear example below). In particular, by using (3.23), we have $(\partial y/\partial x)_{w_0^\partial} = b$ as it has to be, and using (3.20) we obtain indeed $\delta(x, y)$. Then by (3.21) we have $\partial\delta(x, y)/\partial x = \sqrt{2}b/(b+1)$ and $\partial\delta(x, y)/\partial y = -\sqrt{2}/(b+1)$, as it should be.

3.4.2 A linear frontier

We now consider a more general linear model for the frontier when $p, q \geq 1$ and a general direction vector $d = [-d_x' \;\; d_y]'$. The frontier surface is given by the hyperplane in $\mathbb{R}^r$ with equation

$$\gamma_0 + \gamma' w = 0, \tag{3.24}$$

for some $\gamma_0 \in \mathbb{R}$ and $\gamma \in \mathbb{R}^r$.[10] We can derive the distance function, for any $w \in \Psi$,

$$\delta(w) = -\frac{\gamma_0 + \gamma' w}{\gamma' d}. \tag{3.25}$$

Now after simple algebra, it easy to see that the frontier surface in the $t = (v, u)$-space has equation $\gamma_0 + \tilde{\beta}' t = 0$, with $\tilde{\beta} = R_d\gamma$ and we note that $(\tilde{\beta}_1 \ldots \tilde{\beta}_{r-1})' = S_d'\gamma$ and $\tilde{\beta}_r = d'\gamma/||d||$. Equivalently we can write for the frontier points in the t-space:

$$u^\partial = \alpha + \beta' v = \phi(v), \tag{3.26}$$

[10] Note that the parameters are defined at a multiplicative constant $c \neq 0$, i.e. (γ_0, γ) can be replaced by $(c\gamma_0, c\gamma)$ in (3.24) for some $c \neq 0$.

where $\beta = -(\tilde{\beta}_1 \dots \tilde{\beta}_{r-1})'/\tilde{\beta}_r$ and $\alpha = -\gamma_0/\tilde{\beta}_r$. Here we have $\nabla\phi(v) = \beta$.[11]

As we will see in Section 4, we will provide estimates of α and β so the question is how from the knowledge of R_d, $\phi(v)$ and $\nabla\phi(v)$, we can recover all what we need in the $w = (x, y)$-space. From (3.20) we can recover the linear distance function in the $w = (x, y)$-space

$$\delta(w) = ||d||^{-1}\big(\alpha + \tilde{\gamma}'w\big), \text{ where } \tilde{\gamma} = S_d\beta - d/||d||. \tag{3.27}$$

We see that the hyperplane defining the frontier is given by the equation $\delta(w) = 0$, so we can recover $\gamma_0 = ||d||^{-1}\alpha$ and $\gamma = ||d||^{-1}\tilde{\gamma}$ (remember also Footnote 10).

Due to the definitions of α and β above, we have (see also Footnote 9) $\tilde{\gamma} = -\gamma/\tilde{\beta}_r$, the distance function can be written as

$$\delta(w) = -\frac{\gamma_0 + \gamma' w}{||d||\tilde{\beta}_r}, \tag{3.28}$$

Since $\tilde{\beta}_r = d'\gamma/||d||$, we recover the original equation for $\delta(w)$ in (3.25). Note also that from (3.16), we have $c_w = S_d\beta - d/||d|| = \tilde{\gamma}$ and at frontier points, $\partial w_\ell/\partial w_k = -\tilde{\gamma}_k/\tilde{\gamma}_\ell = -\gamma_k/\gamma_\ell$, as it should be.

Finally the derivatives $\partial\delta(w)/\partial w$ can also be obtained from the knowledge of $\phi(v)$. Indeed from (3.21) we have

$$\frac{\partial\delta(w)}{\partial w} = ||d||^{-1}\left(S_d\beta - \frac{d}{||d||}\right) = ||d||^{-1}\tilde{\gamma} = -\frac{\gamma}{d'\gamma}, \tag{3.29}$$

as it should be in our linear model, by (3.25).

3.5 Oriented radial measures

To be exhaustive we summarize here the point of Simar and Wilson (2021) (see SW, Section 3.5.1) showing that the directional model (3.1) can be used for handling radial distance functions. Suppose we are interested in the hyperbolic measures of efficiency (see Färe et al., 1985)

$$\tau(x, y) = \inf\{\tau \mid (\tau x, \tau^{-1}y) \in \Psi\}. \tag{3.30}$$

Here the projection of a point (x, y) on the frontier has coordinates $x^\partial = \tau(x, y)x$ and $y^\partial = \tau^{-1}(x, y)y$. Provided all of the inputs and outputs are strictly positive, we can work with the logs of (X_i, Y_i) and we can write, analogous to (3.1)

$$\begin{bmatrix} \log X_i \\ \log Y_i \end{bmatrix} = \begin{bmatrix} \log X_i^\partial \\ \log Y_i^\partial \end{bmatrix} - \delta_i \begin{bmatrix} -i_p \\ i_q \end{bmatrix}, \tag{3.31}$$

where i_k is a k-vector of ones. Then we obtain $\tau_i = \exp(\delta_i)$. Hence, we are back to our model (3.1), in the log-scale and using the direction vector as in (3.31).

As shown in SW, similar transformations allow to consider the Farrell (1957) radial input efficiency (setting $d = [-i_p' \;\; 0_q']'$ in (3.31)) and the radial output efficiency (here $d = [0_p' \;\; i_q']'$).

4 Best Approximation and Estimation

Hence, the problem is to estimate the function $\phi(\cdot)$ in (3.13) from the sample of iid observations $\{(V_i, U_i)\}_{i=1}^n$, where δ_i has some density, $D_+(\cdot)$, on $\mathbb{R}_+$ with characteristics that may depend on V_i. We repeat here (3.13) for convenience

$$U_i = \phi(V_i) - ||d||\delta_i.$$

This is exactly the paradigm described in Florens and Simar (2005) (hereafter FS) but in the transformed space. We might be tempted to use classical regression techniques to estimate $\phi(\cdot)$, Let $\mu_\delta(v) = \mathbb{E}(\delta(V) \mid V = v)$, we could then use traditional regression techniques for estimating the function $r_1(v) = \phi(v) - ||d||\mu_\delta(v)$ since we have the equation

$$U_i = r_1(V_i) - \varepsilon_i, \tag{4.1}$$

where $\varepsilon_i = ||d||(\delta_i - \mu_\delta(V_i))$ so that now $\mathbb{E}(\varepsilon_i \mid V_i) = 0$. Therefore, least squares techniques may be used (parametric or nonparametric) to provide consistent estimates $\widehat{r}_1(v)$. Then if we fix the particular density $D_+(\cdot)$ for δ_i, we can derive, in most of the cases, the equation of $\mu_\delta(v)$ as a function of its higher moments.

[11] We note that if γ has the same direction than d (i.e., d is orthogonal to the frontier surface), we have $\gamma = kd$ for some $k \neq 0$, then $S_d'\gamma = 0_{r-1}$ and $\beta = 0_{r-1}$, i.e. the surface in the transformed space is, as expected, parallel to the v-axis.

For the one parameter scale family (like Exponential or Half Normal), the knowledge of the variance is enough.[12] This variance may be estimated by regressing in a second stage the squared residuals from the regression in (4.1) on v. Then we can derive $\widehat{\mu}_\delta(v)$ and shift back $\widehat{r}_1(v)$ to get the estimate $\widehat{\phi}(v)$. Simar et al. (2017) have used this technique in the stochastic frontier framework, and it is easy to adapt the method to the deterministic case.

This traditional approach for deterministic models is well known, however, as noted in FS, it suffers from two drawbacks. First, the first stage regression (parametric or nonparametric) to get $\widehat{r}_1(v)$ captures the shape of the cloud of points $\{(V_i, U_i)\}_{i=1}^n$ near its center $(\mathbb{E}(U_i|V_i))$, whereas we want to fit the shape of points near its efficient boundary (U_i^∂). Secondly, we need a parametric family to be able to identify $\mu_\delta(v)$, and the chosen density heavily precondition the characteristic of the final estimate of ϕ. In particular a wrong choice provides wrong estimates.

The method suggested by FS avoids these two drawbacks and can be summarized as follows. First project the observations on a nonparametric frontier and in a second stage, approximate the cloud of estimated frontier points by some suitable parametric model, by using least-squares approximations. FS analyze mainly linear parametric models and show that when using a full nonparametric frontier estimation (like FDH) in the first stage, the obtained estimates converge to the pseudo-true values (the best chosen parametric model to approximate the true frontier). To get inference on the resulting parameters, they use the partial order-m frontiers because they have better rates of convergence and asymptotic normality (Daouia et al., 2008, obtain a similar result for the case of order-α frontiers).

The FS approach adapted to our framework goes along the following lines. Since the frontier function $\phi(v)$ is unknown, we consider a class of parametric functions that can be written as $\{\phi_\theta \mid \theta \in \mathbb{R}^k\}$, where the functions ϕ_θ are defined on $\mathbb{R}^{r-1}$ and depend on a finite number of parameters θ. The best parametric approximation of the true frontier function ϕ in the parametric family $\{\phi_\theta \mid \theta \in \mathbb{R}^k\}$ is defined through the pseudo-true value of θ:

$$\theta_0 = \arg\min_{\theta \in \mathbb{R}^k} \int \big(\phi(v) - \phi_\theta(v)\big)^2 f_V(v)dv. \tag{4.2}$$

If the parametric model is true, this coincides with the true value of θ. As pointed in FS, the existence and uniqueness of the pseudo-true values are based on technical conditions (integrability identification structure of the functional space $\{\phi_\theta \mid \theta \in \mathbb{R}^k\}$). As FS, we consider that this set is squared integrable with respect to $f_V(v)$, then if the set $\{\phi_\theta \mid \theta \in \mathbb{R}^k\}$ is closed and convex, the pseudo true value exists and is unique (see FS for details).

The density f_V is unknown but we can define the "sample" or the "empirical" version of the pseudo-true value by using the empirical discrete density $\hat{f}_{n,V}$, putting a mass $1/n$ at each observed value V_i, $i = 1, \ldots, n$. So we define

$$\begin{aligned} \theta_{0,n} &= \arg\min_{\theta \in \mathbb{R}^k} \sum_{i=1}^n \left[\phi(V_i) - \phi_\theta(V_i)\right]^2, \\ &= \arg\min_{\theta \in \mathbb{R}^k} \sum_{i=1}^n \left[U_i^\partial - \phi_\theta(V_i)\right]^2, \end{aligned} \tag{4.3}$$

since $U_i^\partial = \phi(V_i)$. In practice, U_i^∂ is not observed but we can replace these frontier points by their nonparametric estimators $\widehat{U}_i^\partial$.

So, in our setup, the steps of the method can be described as follows:

[1] From the sample $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$ compute the nonparametric estimators $\widehat{\delta}_i$, $i = 1, \ldots, n$ of the directional distances, and transform the data by the rotation[13]

$$\begin{bmatrix} V_i \\ U_i \end{bmatrix} = R_d \begin{bmatrix} X_i \\ Y_i \end{bmatrix}, \tag{4.4}$$

where R_d is the fixed nonrandom matrix defined in (3.4).

[2] Project the observed U_i on the nonparametric frontier, providing

$$\widehat{U}_i^\partial = U_i + ||d||\widehat{\delta}_i, \tag{4.5}$$

[12] For the Gamma density and for the Truncated normal, we need also the 3rd moment, see Greene (1980) and Stevenson (1980).

[13] We drop the index "•" or "⋆" in $\widehat{\delta}_i$ to indicate which estimator is used (FDH, DEA or partial frontiers).

which are the nonparametric estimates of the unobserved true values

$$U_i^\partial = U_i + ||d||\delta_i. \tag{4.6}$$

[3] Use the sample $\{(V_i, \widehat{U}_i^\partial)\}_{i=1}^n$ to find the best parametric approximate of the function $\phi(\cdot)$, by least squares approximation:

$$\widehat{\theta}_n = \arg\min_{\theta\in\mathbb{R}^k} \sum_{i=1}^{n} \left[\widehat{U}_i^\partial - \phi_\theta(V_i)\right]^2, \tag{4.7}$$

where $\phi_\theta(\cdot)$ is a given class of parametric functions.

The last step provides, for any v, an estimate of the best parametric approximation of the frontier $\widehat{\phi}(v) = \phi_{\widehat{\theta}_n}(v)$ and gives also estimates of its derivatives $\widehat{\nabla\phi}(v) = \widehat{\partial\phi(v)/\partial v}$ and we know from Section 3.3 how to recover, from these estimates, the objects of interest in the original $w = (x, y)$-space.

We analyze in the next section how this particularizes to linear approximations. Then we will investigate how to extend the approach to the more flexible local linear approximations. The properties of the resulting approximations will depend on the chosen nonparametric frontier estimate and on the chosen approximation method.

5 Linear approximation

5.1 The linear model

We start with the simple parametric approximation determined by a linear model.[14] So the model chosen to approximate the frontier in (4.3) and therefore in (4.7) is given by

$$\phi_\theta(v) = \alpha + \beta' v. \tag{5.1}$$

Denote $\theta = [\alpha \quad \beta']'$ the r-vector of parameters. Our least-squares estimator of the best approximation in (4.7) is given by solving

$$(\widehat{\alpha}_n, \widehat{\beta}_n) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} \left[\widehat{U}_i^\partial - (\alpha + \beta' V_i)\right]^2, \tag{5.2}$$

where $\widehat{U}_i^\partial$ is the chosen estimator of the frontier (full or partial) we want to approximate. The solution can be written in closed form

$$\widehat{\theta}_n = \begin{pmatrix} \widehat{\alpha}_n \\ \widehat{\beta}_n \end{pmatrix} = (\boldsymbol{V}'\boldsymbol{V})^{-1}\boldsymbol{V}'\widehat{\boldsymbol{U}}^\partial, \tag{5.3}$$

with the notation $\boldsymbol{V}$ for the augmented data matrix of size $n \times (p+q)$ (augmented by the first column of ones) whose ith row is $[1 \quad V_i']$ and $\widehat{\boldsymbol{U}}^\partial$ is the n-vector of components $\widehat{U}_i^\partial$, $i = 1, \ldots, n$. So we have $\widehat{\phi}(v) = \widehat{\alpha}_n + \widehat{\beta}_n\, v$ and $\widehat{\nabla\phi}(v) = \widehat{\beta}_n$.

5.2 Statistical properties

Florens and Simar (2005) describe the statistical properties of $\widehat{\theta}_n$ and analyze their behavior with respect to the pseudo-true value $\theta_{0,n}$, defined as the solution of the least-squares problem (4.3), providing the best linear approximation of the frontier. Here we have to specify which frontier is approximated. For the full frontier case, the unavailable pseudo-true values are defined as

$$\theta_{0,n} = \begin{pmatrix} \alpha_{0,n} \\ \beta_{0,n} \end{pmatrix} = (\boldsymbol{V}'\boldsymbol{V})^{-1}\boldsymbol{V}'\boldsymbol{U}^\partial, \tag{5.4}$$

[14] For sake of brevity and notational simplicity, we restrict the presentation to the simple linear model in (5.1). As shown in Florens and Simar (2005) and Daouia et al. (2008), all what follows can be extended to linear models of the form $\phi_\theta(v) = g'(v)\theta$, where $g'(v) = (g_1(v) \ldots g_k(v))$ with the k functions $g_j(v)$ being known linearly independent scalar functions of the vector v. See Section 5.3 for an example.

where $\boldsymbol{U}^{\partial}$ is the n-vector of true values $U_i^{\partial} = \phi(V_i)$. Similarly if robust order-m frontiers are used we have

$$\theta_{0,n}^{(m)} = \begin{pmatrix} \alpha_{0,n}^{(m)} \\ \beta_{0,n}^{(m)} \end{pmatrix} = (\boldsymbol{V}'\boldsymbol{V})^{-1}\boldsymbol{V}'\boldsymbol{U}_m^{\partial}, \tag{5.5}$$

where $\boldsymbol{U}_m^{\partial}$ is the n-vector of true order-m values, $U_{m,i}^{\partial} = \phi_m(V_i)$, i.e., $\boldsymbol{U}_m^{\partial} = \boldsymbol{U} + ||d||\boldsymbol{\delta}_m$ where $\boldsymbol{\delta}_m$ is the vector of the n true values of the distance of order-m at the n data points (X_i, Y_i), $i = 1, \ldots, n$. Analog expression would be used for order-α robust frontiers.

5.2.1 Using robust partial frontiers

The advantage of using partial robust frontier (order-α quantile or order-m) is twofold: (i) it provides estimates of nonparametric frontier more robust to outliers and extreme data than traditional envelopment estimators (FDH or DEA) and (ii) the statistical properties of the resulting estimators are much better (functional convergence of the error of estimation with the parametric $\sqrt{n}$-rate to a centered Gaussian process). We will present the arguments for the order-m frontier case as in Florens and Simar (2005), but the presentation would be similar for the order-α case (see Daouia et al., 2008 for details). Here the estimator of the pseudo-true value $\theta_{0,n}^{(m)}$ is given by

$$\widehat{\theta}_n^{(m)} = \begin{pmatrix} \widehat{\alpha}_n \\ \widehat{\beta}_n \end{pmatrix} = (\boldsymbol{V}'\boldsymbol{V})^{-1}\boldsymbol{V}'\widehat{\boldsymbol{U}}_m^{\partial}, \tag{5.6}$$

where $\widehat{\boldsymbol{U}}_m^{\partial} = \boldsymbol{U} + ||d||\widehat{\boldsymbol{\delta}}_{m,n}$, where $\widehat{\boldsymbol{\delta}}_{m,n}$ is the vector of the n estimated values of the distance of order-m at the n data points (X_i, Y_i), $i = 1, \ldots, n$. So we have

$$\begin{aligned} \widehat{\theta}_n^{(m)} - \theta_{0,n}^{(m)} &= (\boldsymbol{V}'\boldsymbol{V})^{-1}\boldsymbol{V}'\big(\widehat{\boldsymbol{U}}_m^{\partial} - \boldsymbol{U}_m^{\partial}\big), \\ &= ||d||(\boldsymbol{V}'\boldsymbol{V})^{-1}\boldsymbol{V}'(\widehat{\boldsymbol{\delta}}_{m,n} - \boldsymbol{\delta}_m). \end{aligned} \tag{5.7}$$

Here the situation is easy, due to the nice asymptotic properties of $\widehat{\boldsymbol{\delta}}_{m,n} - \boldsymbol{\delta}_m$, by using Theorem 3.1 in Florens and Simar (2005), we have as $n \to \infty$

$$\sqrt{n}\Big(\widehat{\theta}_n^{(m)} - \theta_{0,n}^{(m)}\Big) \xrightarrow{\mathcal{L}} N(0, \Sigma), \tag{5.8}$$

where Σ is a matrix depending on several characteristics of the DGP. The structure of the variance of $\sqrt{n}\big(\widehat{\boldsymbol{\delta}}_{m,n} - \boldsymbol{\delta}_m\big)$ is described in Florens and Simar (2005). For practical inference they suggest to estimate Σ by bootstrap methods.[15]

The bootstrap is done in the original units, i.e. generating the bootstrap sample $\mathcal{X}_n^* = \{W_i^* = (X_i^*, Y_i^*), i = 1, \ldots, n\}$ to provide the values (V_i^*, U_i^*) in the (v, u)-space and the bootstrap analog $\widehat{\boldsymbol{\delta}}_{m,n}^*$ of the distances $\widehat{\boldsymbol{\delta}}_{m,n}$. We can use the basic model in (3.1) that describes how the data are generated. So the bootstrap values are defined as $W_i^* = \widehat{W}_i^{\partial} - \delta_i^* d$ where $\widehat{W}_i^{\partial}$ are the projected data points on the nonparametric frontier (DEA or FDH), and δ_i^* are random drawn from some smooth consistent nonparametric estimator of the density of δ, taking into account the boundary condition (see e.g. the details in Section 4.3.5, Simar and Wilson, 2008 or in Section 3.1.2, Simar and Wilson, 2013). The reader can verify that since the W_i^* are generated along the direction d, we have $V_i^* = V_i$ and $U_i^* = \widehat{U}_i - ||d||\delta_i^*$, for $i = 1, \ldots, n$.

By using the sample $\mathcal{X}_n^*$ as reference sample, we can compute the estimator of the order-m distances for all the original data points $\widehat{\boldsymbol{\delta}}_{m,n}^* = \{\widehat{\delta}_{m,n}^*(X_i, Y_i)\}_{i=1}^n$ providing the bootstrap version $\widehat{\boldsymbol{U}}_m^{\partial,*} = \boldsymbol{U} + ||d||\widehat{\boldsymbol{\delta}}_{m,n}^*$. Finally, applying (5.6) in the bootstrap world with $\boldsymbol{V}^* = \boldsymbol{V}$, we obtain $\widehat{\theta}_n^{*,(m)}$ the bootstrap analog of $\widehat{\theta}_n^{(m)}$. By repeating this bootstrap simulation a large number of times, we obtain the bootstrap distribution of $\sqrt{n}\Big(\widehat{\theta}_n^{*,(m)} - \widehat{\theta}_n^{(m)}\Big)$ as approximation of $\sqrt{n}\Big(\widehat{\theta}_n^{(m)} - \theta_{0,n}^{(m)}\Big)$.

The bootstrap also provides the bootstrap values of the quantities of interest described in Section 3.4.2, since they are known (non-random) linear or continuous transformation of θ. From the bootstrap distribution, we can, e.g., evaluate confidence intervals for these objects. For all these quantities the basic bootstrap method is recommended (rather than the percentile method) due to the possible bias in finite samples for these quantities.

[15] For the order-α case, we refer to Theorem 3.1 in Daouia et al. (2008).

5.2.2 Using full frontiers

For the full frontier we can use either the FDH or the DEA estimates $\widehat{\delta}_{n,i}$.[16] Here the story is less easy due to the bias introduced by the DEA/FDH estimators and the lack of functional convergence of these estimators. Here we have, analog to (5.7), the relation

$$\widehat{\theta}_n - \theta_{0,n} = ||d||(\boldsymbol{V}'\boldsymbol{V})^{-1}\boldsymbol{V}'(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}), \tag{5.9}$$

where $\widehat{\boldsymbol{\delta}}_n$ (and $\boldsymbol{\delta}$) is the vector of the n nonparametric estimators (resp. the n true values) of the efficiency scores evaluated at the data points (X_i, Y_i), $i = 1, \dots, n$.

Theorem 3.2 in Florens and Simar (2005) implies that $\widehat{\theta}_n - \theta_{0,n} \xrightarrow{p} 0$, which is fine but not enough for practical inference on $\theta_{0,m}$. Equation (5.9) indicates that the rates of convergence will be the same as the original FDH/DEA estimator used (see Kneip et al., 2015).

5.3 Cobb-Douglas and Translog cases

Some practitioners like to use in the original w-space generalized Cobb-Douglas or Translog parametric models for the frontier. These are particular cases of linear models (linear in the parameters) but in the log-scale.

The Cobb-Douglas is a first order approximation in the log-scale and since the rotation is a linear transformation, the link between the two spaces is immediate. Denote by $\widetilde{W}_i$ the original data and so $W_i = \log \widetilde{W}_i$. We can then follow what is described above keeping the same notations. In the units w, we have the linear model given in (3.24)

$$\gamma_0 + \gamma' w = 0, \tag{5.10}$$

and after the rotation R_d we obtain the linear model described in (3.26) which is indeed linear in $t = R_d w$ and has the linear form given in (5.1). Then from the observed values of $V_i, \widehat{U}_i^{\partial}$ we obtain the estimates $\widehat{\alpha}_n$ and $\widehat{\beta}_n$.

Hence, to get the estimates of the parameters in the generalized Cobb-Douglas model in the original units w, we have $\widehat{\widetilde{\gamma}}_n = S_d\widehat{\beta}_n - d|||d||$, which is all we need to characterize the frontier, see Section 3.4.2. Indeed from (3.27), the frontier function can be written for any $w = log(\widetilde{w})$ as

$$\widehat{\alpha}_n + \widehat{\widetilde{\gamma}}_n' w = 0. \tag{5.11}$$

The Translog idea is based on second order approximation, therefore it is a bit more complicated and the links are not so immediate. If we use a quadratic model in the transformed space, instead of (5.1) we have

$$\phi(v) = \alpha + \beta' v + \frac{1}{2} v' \Lambda v, \tag{5.12}$$

where $\Lambda' = \Lambda$. This model can be viewed as an extension of the linear model since it allows quadratic terms in v. The linear model corresponds to the case $\Lambda = 0$. This model can be estimated by usual least-squares techniques (adapting the notation in (5.3)) providing estimates of (α, β, Λ). Then by using (3.22), we obtain for any $w = log(\widetilde{w})$ on the frontier

$$\alpha + \gamma' w + \frac{1}{2} w' \Gamma w = 0, \tag{5.13}$$

where $\gamma = S_d\beta - d|||d||$ and $\Gamma = S_d \Lambda S_d'$, fixing the parameters of the frontier in the units w. This looks like a Generalized Translog, but it is a "constrained" translog model. Indeed, it is easy to see that, since $S_d' d = 0_{r-1}$, we have $\Gamma d = 0_r$, which remind us that the second order terms have been introduced in the v-space which is orthogonal to d.

Note that in the particular case mentioned at the beginning of Section 3.4, when $q = 1$, $p \geq 1$, $d_x = 0_p$ and $d_y = 1$, there is no need for a rotation: $R_d = I_r$, $u = y$ and $v = x$ and we have the usual Translog model for a production frontier.

[16]The DEA estimators provides estimates with a better rate of convergence, but they are not consistent if the set Ψ is not convex.

6 Local linear approximation

When it is difficult to specify a priori a global parametric model for the frontier function ϕ, using more flexible local parametric approximation would allow us a richer interpretation of its shape providing also its local derivatives. We investigate now how local linear models can easily be adapted to our setup.

In place of looking for the best linear approximation, we might indeed search for more flexible approximations for functions $\phi(v)$ which admit for all values of v a local linear approximation. If the true function $\phi(v)$ is smooth enough (differentiable till order 2), we can use the first terms of a Taylor expansion of the function around v

$$\phi(\tilde{v}) = \phi(v) + \left(\frac{\partial\phi(\tilde{v})}{\partial v}\right)'_{\tilde{v}=v} (\tilde{v} - v) + o(||\tilde{v} - v||), \tag{6.1}$$

and the leading terms can be written, for $\tilde{v}$ in a neighborhood of v as

$$\phi(\tilde{v}) = \alpha(v) + \beta'(v)(\tilde{v} - v). \tag{6.2}$$

Here, is the spirit of (4.2), the pseudo-true values are defined as the best local linear approximation of $\phi(v)$, so we can define the local pseudo-true value at any point v as

$$(\alpha_0(v), \beta_0(v)) = \arg\min_{\alpha,\beta} \int \left(\phi(\tilde{v}) - [\alpha + \beta'(\tilde{v} - v)]\right)^2 K_h(\tilde{v} - v) f_V(\tilde{v}) d\tilde{v}, \tag{6.3}$$

where $K_h(\tilde{v} - v)$ is a weighting function localizing the values $\tilde{v}$ in a neighborhood of v. We can use any standard multivariate kernel function and h is a bandwidth vector tuning the weights. In practice we will use a product kernel, so that

$$K_h(\tilde{v} - v) = \prod_{j=1}^{r-1} \frac{1}{h_j} K\big(\frac{\tilde{v}_j - v_j}{h_j}\big), \tag{6.4}$$

and $K(\cdot)$ is a simple univariate kernel function. We will use below kernels with compact support, i.e $K(u) = 0$ when $|u| > 1$, like e.g. Epanechnikov kernels. The empirical version of the pseudo-true values are defined similarly to (4.3) as

$$(\alpha_{0,n}(v), \beta_{0,n}(v)) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} \left(U_i^\partial - [\alpha + \beta'(V_i - v)]\right)^2 K_h(V_i - v), \tag{6.5}$$

where as above $U_i^\partial = \phi(V_i)$ are the true (unobserved) frontier points. The nice thing of local linear approximations, is that a closed form is available for the solution in (6.5). It is known (see e.g. Fan and Gijbels, 1996) that

$$\theta_{0,n}(v) = \begin{pmatrix} \alpha_{0,n}(v) \\ \beta_{0,n}(v) \end{pmatrix} = (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{U}^\partial, \tag{6.6}$$

where the $n \times (p+q)$ matrix $\boldsymbol{\mathcal{V}}$ has its ith row given by $[1 \quad (V_i - v)']$ and the $n \times n$ weights matrix $\boldsymbol{W}(v)$ is diagonal with ith element given by $K_h(V_i - v)$ and $\boldsymbol{U}^\partial$ is the n-vector of true values of frontier points, i.e. $\phi(V_i)$.

For the order-m frontier, we have

$$\theta_{0,n}^{(m)}(v) = \begin{pmatrix} \alpha_{0,n}^{(m)}(v) \\ \beta_{0,n}^{(m)}(v) \end{pmatrix} = (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{U}_m^\partial, \tag{6.7}$$

where now $\boldsymbol{U}_m^\partial$ is the n-vector of true values of order-m frontier points, i.e. $\phi_m(V_i)$.

If the frontier functions are sufficiently smooth (differentiable till order 2) we know by (6.1) that $\alpha_{0,n}(v) \to \phi(v)$ and $\beta_{0,n}(v) \to \partial\phi(\tilde{v})/\partial v$ and that $\alpha_{0,n}^{(m)}(v) \to \phi_m(v)$ and $\beta_{0,n}^{(m)}(v) \to \partial\phi_m(\tilde{v})/\partial v$ as $h \to 0$.

Now the true values $\boldsymbol{U}^\partial$ are unavailable, but as above we will in practice use their appropriate (FDH/DEA or robust) estimators $\widehat{\boldsymbol{U}}^\partial$. So the (local) values of $\theta(v) = [\alpha(v) \quad \beta'(v)]'$ will be estimated from the sample $\{V_i, \widehat{U}_i^\partial\}_{i=1}^n$ by the weighted constrained least squares problem

$$(\widehat{\alpha}_n(v), \widehat{\beta}_n(v)) = \arg\min_{\alpha,\beta} \left[\sum_{i=1}^{n} \left(\widehat{U}_i^\partial - (\alpha + \beta'(V_i - v))\right)^2 K_h(V_i - v)\right]. \tag{6.8}$$

From the Taylor expansion (6.1), it is clear that $\widehat{\alpha}_n(v)$ is the estimated smoothed value of $\phi(v)$ and that $\widehat{\beta}_n(v)$ provides an estimator of the first derivatives $\nabla\phi(v)$. They are computed by

$$\widehat{\theta}_n(v) = \begin{pmatrix} \widehat{\alpha}_n(v) \\ \widehat{\beta}_n(v) \end{pmatrix} = (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\widehat{\boldsymbol{U}}^{\partial}. \tag{6.9}$$

Clearly we have for all v,

$$\begin{aligned} \widehat{\theta}_n(v) - \theta_{0,n}(v) &= (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\big(\widehat{\boldsymbol{U}}^{\partial} - \boldsymbol{U}^{\partial}\big) \\ &= ||d||(\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\big(\widehat{\boldsymbol{\delta}}^{\partial} - \boldsymbol{\delta}^{\partial}\big), \end{aligned} \tag{6.10}$$

which is similar to what we have above for the simple linear approximations in (5.9). In practice the bandwidths h are determined by least-squares cross validation (LSCV) and this provides bandwidths with an optimal order $h_j = O\big(n^{-1/(r+3)}\big)$ since $v \in \mathbb{R}^{r-1}$. Since for a given v, (6.10) is a simple linear transformation of the estimation errors $\widehat{\boldsymbol{\delta}}^{\partial} - \boldsymbol{\delta}^{\partial}$, we keep the same properties as described in Florens ans Simar (2005), i.e. only consistency for full frontier approximations, i.e. for all v we have $\widehat{\theta}_n(v) - \theta_{0,n}(v) \xrightarrow{p} 0$ as $n \to \infty$.

For the order-m frontiers, we have

$$\widehat{\theta}_n^{(m)}(v) = \begin{pmatrix} \widehat{\alpha}_n^{(m)}(v) \\ \widehat{\beta}_n^{(m)}(v) \end{pmatrix} = (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\widehat{\boldsymbol{U}}^{\partial}_m. \tag{6.11}$$

Clearly we have for all v, as $n \to \infty$

$$\begin{aligned} \widehat{\theta}_n^{(m)}(v) - \theta_{0,n}^{(m)}(v) &= (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\big(\widehat{\boldsymbol{U}}^{\partial}_m - \boldsymbol{U}^{\partial}_m\big) \\ &= ||d||(\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\big(\widehat{\boldsymbol{\delta}}^{\partial}_m - \boldsymbol{\delta}^{\partial}_m\big), \end{aligned} \tag{6.12}$$

where here again the bandwidths can be selected by LSCV. So for a given v we keep the proprties established in Theorem 3.1 of Florens and Simar (2005), specifically, for any v,

$$\sqrt{n}\Big(\widehat{\theta}_n^{(m)}(v) - \theta_{0,n}^{(m)(v)}\Big) \xrightarrow{\mathcal{L}} N(0, \Sigma(v)), \tag{6.13}$$

where $\Sigma(v)$ is a matrix depending on several characteristics of the DGP. Here again practical inference will be obtained with the bootstrap as described above for the simple linear case.

To summarize, we generate the bootstrap sample according to the basic model, i.e. $W_i^* = \widehat{W}_i - \delta_i^* d$, which provides the values $V_i^* = V_i$ and $U_i^* = \widehat{U}_i - ||d||\delta_i^*$, for $i = 1, \ldots, n$. We have also, as for the linear case above, the estimator of the order-m distances for all the original data points $\widehat{\boldsymbol{\delta}}^*_{m,n} = \{\widehat{\delta}^*_{m,n}(X_i, Y_i)\}_{i=1}^n$ providing the bootstrap version $\widehat{\boldsymbol{U}}^{\partial,*}_m$ of $\widehat{\boldsymbol{U}}^{\partial}_m$. Finally applying (6.11) in the bootstrap world, keeping the same matrices $\boldsymbol{V}$ and $\boldsymbol{W}(v)$, we obtain the value $\widehat{\theta}_n^{*,(m)}(v)$, the bootstrap analog of $\widehat{\theta}_n^{(m)}(v)$ which in turn can provide, by Monte-Carlo simulations, the bootstrap approximation to (6.13) (see above Section 5.2.1 for additional details).

7 Dealing with environmental factors

In the presence of environmental factors Z, which are neither inputs nor outputs, but are factors that might influence the production process, Cazals et al. (2002) and Daraio and Simar (2005) have introduced he concept of conditional frontiers and conditional efficiency measures. This leads to define a conditional attainable set

$$\Psi^z = \{(x, y) \mid x \text{ can produce } y \text{ when } Z = z\}. \tag{7.1}$$

Clearly $\Psi^z \subset \Psi$ which includes as particular case $\Psi^z = \Psi$ for all z. The latter is known as the "separability" condition, which may be quite restrictive in many applications (see Simar and Wilson, 2007). A formal test of separability has been derived in Daraio et al. (2018) and Simar and Wilson (2020). If the model is non-separable, the unconditional efficiency measures have no real economic meaning since they describe the distance for the unit (x, y) to the boundary of Ψ instead of Ψ^z if the unit faces the condition z for Z. So, in this case it is more appropriate to define conditional efficiency measures as

$$\delta(x, y \mid z) = \sup\{\delta \mid (x - \delta d_x, y + \delta d_y) \in \Psi^z\}. \tag{7.2}$$

Nonparametric estimators have been proposed by Cazals et al. (2002) and Daraio and Simar (2005) and most of its statistical properties are derived in Jeong et al. (2010). CLTs have been obtained for averages of these measures in Daraio et al. (2018).

If Z is separable, it has no effect on the frontier of the attainable set, so z has no influence on the shape of the frontier and on the quantities of interest developed in this paper. So the analysis for the full frontier models above can be completed without reference to Z, by using the unconditional measures.

If Z is non-separable or if we are interested to partial frontiers (because, in any case, Z may influence the partial frontier levels), then the measures $\delta(x, y)$ above should be replaced by the conditional measures $\delta(x, y \mid z)$ and their estimators. The transformation of w to t in (3.5) remains the same providing, by (3.9), the transformed data. Then in the approximation in the (v, u) space developed above, we could introduce the additional variables z to approximate the frontiers. Hence we would have in place of (4.5)

$$\widehat{U}^{\partial}(Z_i) = U_i + ||d||\widehat{\delta}(W_i \mid Z_i), \tag{7.3}$$

and the approximating equation (4.7) becomes

$$\widehat{\phi}(\cdot,\cdot) = \arg\min_{\phi(\cdot,\cdot)} \sum_{i=1}^{n} \left[\widehat{U}^{\partial}(Z_i) - \phi(V_i, Z_i)\right]^2, \tag{7.4}$$

where $\phi(\cdot,\cdot)$ belongs to the class of linear or local linear models, as above. Note that here, even with partial frontiers, the dimension of Z introduces some curse of dimensionality, because the estimators of the conditional measures have convergence rates deteriorated by the dimension of Z, for the order-m and order-α, $\sqrt{n}$ becomes $n^{2/(r_z+4)}$, where r_z is the dimension of Z.

Note also that the bootstrap method for the order-m case has to be slightly modified to handle the possible dependence on Z. We generate the bootstrap sample on the "pairs", i.e. here on (W_i, Z_i) to keep the dependence between (X, Y) and Z in the bootstrap sample, providing the bootstrap sample $\{(W_i^*, Z_i^*)\}_{i=1}^n$. The latter sample gives the bootstrap analog $\widehat{\delta}_m^*(W_i \mid Z_i)$ of $\widehat{\delta}_m(W_i \mid Z_i)$, evaluated at the original data points (W_i, Z_i). Note that here the bootstrap analog of $\widehat{U}_m^{\partial}(Z_i)$ will be defined as

$$\widehat{U}_m^{\partial,*}(Z_i) = U_i + ||d||\widehat{\delta}_m^*(W_i \mid Z_i). \tag{7.5}$$

Apart from this, the procedure goes as above for the unconditional cases. The procedure will be illustrated in a real data example below.

8 Numerical Illustrations

8.1 Some simulated examples

8.1.1 A linear frontier and linear approximation

Many researchers assume that the unknown frontier is linear (after eventually a log-transformation). We can handle this case for the directional distance case and avoiding the drawbacks of the traditional regression type methods and avoiding also to specify the distribution of the efficiencies. We illustrate this for the case of one output ($q = 1$) and two inputs ($p = 2$) but, which is unusual, we use a directional distance with $d = (-1 \quad 1)'$. Note that even here, in this simple example, traditional parametric methods cannot be used without our transformation (rotation) due to our chosen $d = (-1 \quad 1)'$. Our frontier model is given by

$$y^{\partial} = 1 + 2x_1^{\partial} + 2x_2^{\partial}. \tag{8.1}$$

According to our DGP in Section 3.1 we generate points $(X_i^{\partial}, Y_i^{\partial})_{i=1}^n$ on the frontier. We first simulate the two components of X_i^{∂} by two independent uniform on $(0, 1)$, i.e. $X_{i,k}^{\partial} \sim \text{Unif}(0, 1)$ for $k = 1, 2$. Hence the Y_i^{∂} are given by (8.1). Finally the observations $(X_i, Y_i)_{i=1}^n$ are generated according to our DGP in (3.1), i.e.

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{bmatrix} X_i^{\partial} \\ Y_i^{\partial} \end{bmatrix} - \delta_i \begin{bmatrix} -d_x \\ d_y \end{bmatrix},$$

where the δ_i are independent half-normal distributed, $\delta_i \sim N^+(0, \sigma_\delta^2)$. For the illustration here we select $n = 200$ and $\sigma_\delta = 0.20$.

By using the results of Section 3.4.2, we can compute the true values of the full frontier equation in the transformed space $t = (v, u)$. For instance the full frontier is determined by the equation

$$u^\partial = \alpha + \beta v.$$

Note that we do not have the true equation of the order-m frontier, but we use here the order-m frontier as a robust estimate of the frontier, so we select for illustration $m = 50, 100$ and $m = 200$, to give some insights on the sensitivity of the robust order-m frontiers to the choice of m. Table 1 provides the true values of the estimates with the full frontier (FDH) projections and with the order-m values. Of course

Table 1: Simulated example, linear case with $n = 200$, frontier in the (v, u) space.

	True	FDH	$m = 50$	$m = 100$	$m = 200$
α	0.3464	0.2270	0.1342	0.1745	0.2019
β_1	0.0732	0.0558	0.0462	0.0491	0.0524
β_1	0.2732	0.2745	0.2573	0.2649	0.2703

we are not so much interested to the values in the (v, u) space, but as explained above we can recover the estimate of the frontier in the original space. We standardise the γ's such that the coefficient of y, $\gamma_3 = -1$ to recover the values in (8.1) (see Footnote 10). The results are given in Table 2. We see that the order-m values approaches the FDH values. Of course, by construction (being robust against potential outliers) the m-frontiers are "below" the FDH frontier, as seen by the values of the intercept, note that for the latter, from $m = 100$, both estimates are quite similar (as they should when $m \to \infty$). By using the results from

Table 2: Simulated example, linear case with $n = 200$, frontier in the (x, y) space.

	True	FDH	$m = 50$	$m = 100$	$m = 200$
γ_0	1.0000	0.6502	0.3681	0.4874	0.5717
γ_1	2.0000	2.0157	1.9174	1.9605	1.9915
γ_2	2.0000	1.9463	1.8323	1.8770	1.9139

Section 3.4.2 we can also estimate the partial derivatives of the distance function $\delta(x, y)$ with respect to all the inputs and the output. Here, in our simulated linear model, the true values are constant (not dependent of (x, y)) and they are given by (3.29). The results are displayed in Table 3. We see that these partial derivatives are very well estimated, even from the order-m estimates as small as $m = 50$.

Table 3: Simulated example, linear case with $n = 200$, derivatives of $\delta(x, y)$.

	True	FDH	$m = 50$	$m = 100$	$m = 200$
$\partial\delta(x, y)/\partial x_1$	0.4000	0.4062	0.4037	0.4053	0.4060
$\partial\delta(x, y)/\partial x_2$	0.4000	0.3923	0.3858	0.3880	0.3902
$\partial\delta(x, y)/\partial y$	-0.2000	-0.2015	-0.2105	-0.2067	-0.2039

Finally we show how the bootstrap works for the parameters of the frontier in the $w = (x, y)$ space and also for the partial derivatives $\partial\delta(w)/\partial w_k$. We use the basic-bootstrap techniques with $B = 1000$ loops and we provide in the tables the 95% confidence intervals. In Table 4 the confidence intervals are for the pseudo-true value $\theta_{0,n}^{(m)}$ of the order-m frontier $\phi_m(x, y)$ (that are unknown). We have seen in (6.13) that these intervals have the parametric $\sqrt{n}$ precision.

If we want the confidence intervals for $\theta_{0,n}$ the parameters of the full frontier $\phi(x, y)$, we cannot use the FDH estimators since we don't have the asymptotic distribution for the full frontier case. But we can use the confidence intervals of Table 4 if m is large enough. In addition these estimates are robust to outliers. We show how the confidence intervals (obtained from one sample) cover the true values of $\theta_{0,n}$ (given in the first column). As a matter of fact, only the intercept is less well estimated, whereas for the most important parameters, i.e. the slopes γ_1 and γ_2, the robust order-m approximations provide useful information. This was already noticed in Florens and Simar (2005) and this is due to the inherent bias of using order-m frontiers to estimate the full frontier.[17] Table 5 provides another important information for the practitioner:

Table 4: Simulated example, linear case with $n = 200$, 95% bootstrap confidence intervals for frontier parameters of $\phi_m(x, y)$.

	True values	$m = 50$		$m = 100$		$m = 200$	
γ_0	1.0000	0.4192	0.6511	0.5374	0.7819	0.6233	0.8819
γ_1	2.0000	1.7736	2.0288	1.8271	2.0955	1.8583	2.1379
γ_2	2.0000	1.7203	1.9581	1.7591	2.0052	1.8042	2.0547

confidence intervals for the partial derivatives of $\delta_m(x, y)$. Here we see indeed that from $m = 100$ we have even the relevant information for $\delta(x, y)$. We observe also that all the derivatives have the appropriate sign.

Table 5: Simulated example, linear case with $n = 200$, 95% bootstrap confidence intervals for derivatives of $\delta_m(x, y)$.

	True values	$m = 50$		$m = 100$		$m = 200$	
$\partial\delta(x, y)/\partial x_1$	0.4000	0.3817	0.4202	0.3848	0.4250	0.3861	0.4258
$\partial\delta(x, y)/\partial x_2$	0.4000	0.3690	0.4072	0.3683	0.4082	0.3716	0.4099
$\partial\delta(x, y)/\partial y$	-0.2000	-0.2178	-0.2033	-0.2196	-0.1981	-0.2106	-0.1945

8.1.2 A nonlinear frontier and its local linear approximation

We analyze here a more complicated situation where the frontier model is nonlinear and we will use the local linear approximations. We simulate the frontier points in a multivariate setup, inspired by the DGP defined in Daraio et al. (2018). We first generate a $(p+q)$-tuple $s = (s_p', s_q')'$ uniformly distributed on a unit sphere centered at the origin of $\mathbb{R}^{p+q}$, where s_p and s_q are column vectors of length p and q, respectively. Then we set $X^\partial = 2(1 - |s_p|)$ and $Y^\partial = 2|s_q|)$. We repeat this for $i = 1, \ldots, n$ to obtain $W_i^\partial = (X_i^\partial, Y_i^\partial)$. The observations $W_i = (X_i, Y_i)$ are generated according to our DGP in (3.1), i.e. $W_i = W_i^\partial - \delta_i d$, where $d = (-i_p', i_q')'$ (i_k denotes a k-vector of ones) and δ_i are independent half-normal distributed, $\delta_i \sim N^+(0, \sigma_\delta^2)$ with $\sigma_\delta = 0.10$. In the example below we will choose $p = q = 2$.

For illustration, Figure 1 displays the frontier points and the data in the simple case where $p = q = 1$ both in the original (x, y) space (left panel) and in the transformed (v, u) space (right panel). For the estimation we use with $p = q = 2$, $n = 400$ observations and we fix $m = 200$. The bandwidths obtained by LSCV for the full frontier case, using the FDH estimators $\widehat{U}_i^\partial$, are $h_{FDH} = (0.6204, 0.4959, 0.3442)'$. Here in this simulation scenario we have the true values of the frontier points U_i^∂, so we can compute the pseudo-true values $\theta_{0,n}(V_i)$, by (6.6). Figure 2 allows to see that, as expected, the pseudo-true values of the frontier fit $\alpha_{0,n}(V_i)$ are not far from the true values U_i^∂. For the order-m frontier, the LSCV provides $h_m = (0.4262, 0.5067, 0.3423)'$. Here we do not have the true values U_m^∂, so we cannot evaluate the pseudo-true values $\theta_{0,n}^{(m)}(V_i)$, however, since $m = 200$ is large we can compare the estimated values $\widehat{\alpha}_n^m(V_i)$, computed by (6.11) with the FDH values $\widehat{\alpha}_n(V_i)$ and with the pseudo-true full frontier values $\alpha_{0,n}(V_i)$. This is shown in Figure 3 to appreciate if the three distributions of the n values of the frontier seem to be coherent. As we noticed above for the simple linear case, the distributions are similar with a bias due to the FDH and the bias of the order-m relative to the full frontier values. To see what is the effect of this when estimating the quantities of economic interest, we first look at some of the marginal rates $\partial y_\ell/\partial x_k$. By using equations (3.16) and (3.17) we compute those for $k = 1$. Figure 4 displays the boxplots of the distribution of these values over the n points, for various estimators, including the true full frontier values and the pseudo-true values for the full frontiers. It is easy to see that the true values are given by $\partial y_\ell/\partial x_k = (2 - x_k)/y_\ell$. Of course the same could be done for x_2. We see that globally the estimators (both the FDH and the order-m cases) have a similar distribution than the true values and of course of the pseudo-true values. Again we do not have the order-m pseudo-true values

[17] Daouia et al. (2012) investigate how order-m frontier estimators can be tuned to estimate the full frontier, by letting $m = m(n) \to \infty$ when $n \to \infty$. The application of these concepts to our setup is rather complicated and is left for future research.

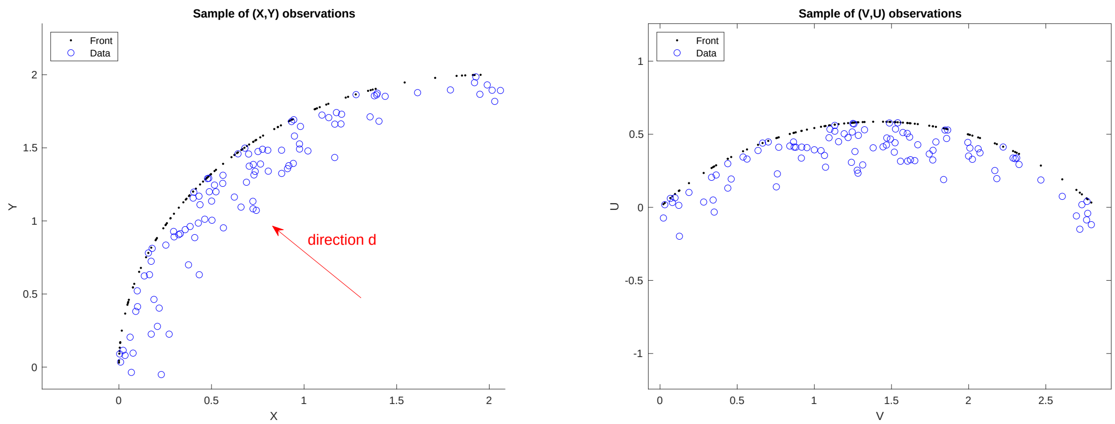

Figure 1: Nonlinear case: a simulated data set of size $n = 100$ with $p = q = 1$.

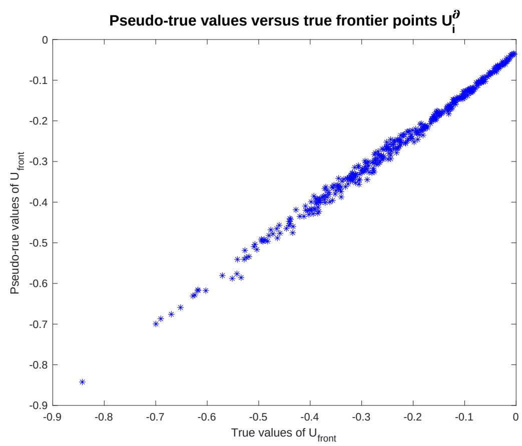

Figure 2: Nonlinear case: pseudo-true fitted values $\alpha_{0,n}(V_i)$ versus true values U_i^{∂}, $n = 400$ with $p = q = 2$.

but we see that with $m = 200$ we have rather good estimates of the distribution of the marginal rates over the sample data points. We note that most of the estimated ratios, evaluated at data points, have the expected right sign, except a few cases when using the FDH estimator. Finally, we evaluated how the estimators of the partial derivatives of $\delta(x, y)$ with respect to the $p+q$ elements (x, y) behaves. They are given by (3.21), where $\partial\phi(v)/\partial v$ is evaluated for each value of V_i by using the corresponding values of $\beta(v)$ obtained by our local linear approximations: i.e. the pseudo-true values $\beta_{0,n}(v)$, the FDH estimators $\widehat{\beta}_n(v)$ and the order-m estimators $\widehat{\beta}_n^{(m)}(v)$ given by (6.6), (6.9) and (6.11) respectively. In the latter cases, the estimates provide estimation of the partial derivatives of $\delta_m(x, y)$, but with large values of m, we expect that $\delta_m(x, y) \approx \delta(x, y)$. We can observe how well the distribution of the estimators follows the distribution of the pseudo-true values for the full frontier, even for the order-m case with $m = 200$. We observe also that most of the derivatives evaluated at the data points have the expected right sign. We can also, if of economic interest, look at the derivatives of the distance function with respect to the variables, i.e., $\partial\delta(w)/\partial w_k$ as a function of w_k. We show below in Figure 6, just for illustration, the cases for $\partial\delta_m(x, y)/\partial x_1$ and $\partial\delta_m(x, y)/\partial y_1$.

Now we estimate the sampling distribution of the order-m estimators by using the bootstrap. We are using $B = 1000$ bootstrap loops and we can provide, for each quantity of interest estimated above, bootstrap confidence intervals. To illustate this we will provide confidence intervals for the partial derivatives evaluated at each data points. Table 6 displays the 95% confidence intervals obtained by the basic-bootstrap method of the partial derivatives $\partial\delta_m(x, y)/\partial x$ and $\partial\delta_m(x, y)/\partial y$. We limit the presentation of this simulated example to the first 20 data points to save space.

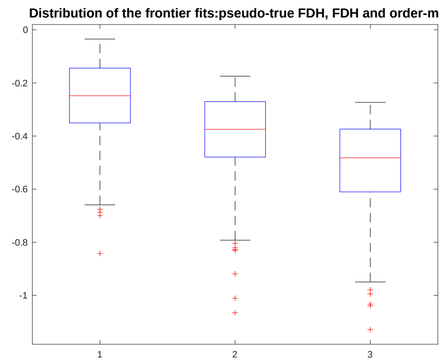

Figure 3: Nonlinear case: distribution of the fitted frontier values, from left to right, $\alpha_{0,n}(V_i)$ (boxplot 1), $\widehat{\alpha}_n(V_i)$ (boxplot 2) and $\widehat{\alpha}_n^m(V_i)$ ((boxplot 3)), $i = 1, \ldots, n$. Here $n = 400$, $m = 200$ with $p = q = 2$.

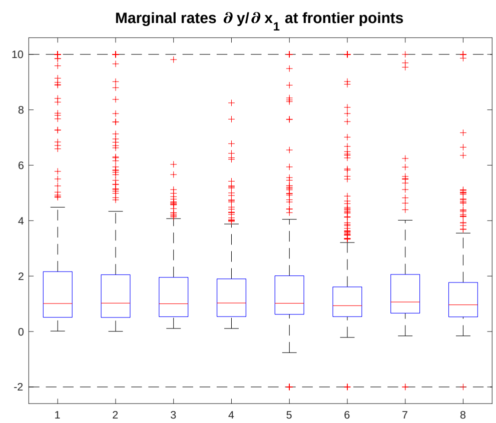

Figure 4: Nonlinear case: distribution of the marginal rates, from left to right by pairs: $(\partial y_1/\partial x_1, \partial y_2/\partial x_1)$, the true values (cases 1 and 2), the full frontier pseudo-true values (cases 3 and 4), the FDH estimators (cases 5 and 6) and the order-m estimators (cases 7 and 8). Here $n = 400$, $m = 200$ with $p = q = 2$.

We see that these individual confidence intervals are rather narrows, showing that the estimation of the partial derivatives of the order-m distance function are quite reliable. This is not a surprise since we have by (6.13) the $\sqrt{n}$-parametric rate of convergence. We see also that most of the intervals cover values with expected sign and in the Table, only the unit #15 has an interval covering the value zero.

Hence globally, this nonlinear example indicates that the local linear approximations can be handled to smooth nonparametric frontiers (here the FDH and the order-m) leading to reasonable estimates of the quantities of economic interest. In addition the bootstrap provides an easy way to make reliable inference by using the robust order-m estimators.

8.2 Application on European higher education data

Efficiency analysis in education has a long tradition (see, e.g., Ruggiero, 2004 and Johnes, 2006). The analysis of European universities is more recent and has been developed from the pioneer project called AQUAMETH described in Bonaccorsi and Daraio (2007), in which the first empirical evidence at the European comparative level is also reported (see also Bonaccorsi et al., 2007). A number of recent surveys have shown a steady increase in the quantity and variety of contributions proposed to assess the efficiency of education in general and higher education in particular (see, e.g., Grosskopf et al. 2014, and De Witte and Lopez-Torres, 2017).

In this section we illustrate our methodology by analyzing a sample of data from European universities

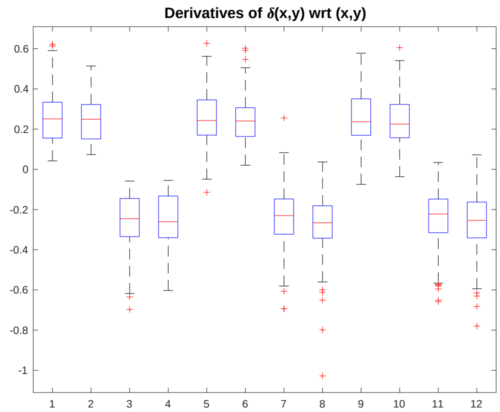

Figure 5: Nonlinear case: distributions of the partial derivatives of $\delta(x, y)$ with respect to the $p + q$ elements (x, y), they are given from the left to the right by quadruples $(\partial\delta(x, y)/\partial x_1, \partial\delta(x, y)/\partial x_2, \partial\delta(x, y)/\partial y_1, \partial\delta(x, y)/\partial y_2)$: cases 1 to 4, pseudo-true values, cases 5 to 8, the FDH estimators and cases 9 to 12, the order-m estimators. Here $n = 400$, $m = 200$ with $p = q = 2$.

that was recently analysed in Daraio et al. (2021) to which readers are referred for more details and information.

Data include as inputs total number of academic staff (ACAD), total number of non-academic staff (NONAC) and total expenditures (TEXP) that is the sum of all expenditures (includes expenditure for personnel, non-personnel, capital and unclassified expenditure); as outputs total number of degrees (TDEG) in all the educational levels without the PhDs which are considered as an additional output (PHD) and total number of publications (PUB) considered as an output which includes the total number of scholarly publications indexed in Scopus. As external-environmental variables we consider the quality latent factor (QUAL) identified in Daraio et al. (2021) and the specialization index (SPEC) that measures the thematic concentration/dispersion of scientific outputs, varying between 0 and 1 and indicating, respectively, generalist versus specialist universities.

Due to the high correlation between the three inputs and between the two research outputs (PUB and PHD), Daraio et al. (2021) applied a dimension reduction analysis based on factor analysis, as suggested in Daraio and Simar (2007) and analyzed by Monte-Carlo analysis in Wilson (2018). The resulting input factor, denoted FX, is determined by the first eigenvector of the second moment matrix of the three inputs $u_x = (0.5723\ 0.6218\ 0.5346)'$, which can roughly be interpreted as an average of the scaled inputs; it explains 96% of the total inertia and so little information is lost by using this single input factor. Its correlations with the three original inputs are 0.9777, 0.9474 and 0.9325 respectively. For the two research outputs we have similar results with $u_y = (0.6986\ 0.7155)'$ which explains 97% of the total inertia. The resulting output research factor, denoted FY, has correlations 0.9676 and 0.9691 with PUB and PHD, respectively. So we end up with 337 observations with one input $X = FX$ and two outputs $Y = (TDEG, FY)$ the first one being the teaching activity and the second summarizing the research activity. A descriptive statistics on the sample is illustrated in the following table taken from Daraio et al. (2021).

In the analysis, all the variables are scaled by their empirical standard deviation. This improves the numerical stability when selecting the optimal bandwidths. So, all the derivatives have to be rescaled by $scX = 1.6471$, $scY = (3196.67,\ 0.001369)$ and $scZ = (0.2903,\ 0.1249)$.

We test the separability condition according to Daraio et al. (2018) and Simar and Wilson (2020) obtaining a p-value near zero, hence we reject the separability condition and we work with conditional frontiers. The sensitivity analysis carried out for selecting the value of m to compute order$-m$ measures showed an elbow effect around a value of $m = 550$, for which 31% of observations lies above the (marginal) order-m frontier, showing negative values of $\widehat{\delta}_m(X_i, Y_i)$. For the conditional order-m frontiers, that include also the external factors QUAL and SPEC, the elbow effect is also present around $m = 550$ but with only 10% of the points located above the frontier, with negative values of $\widehat{\delta}_m(X_i, Y_i|Z_i)$.

The optimal bandwidths h_z for conditional FDH and conditional order-m frontiers, estimated by LSCV are respectively $h_z = (0.9099,\ 0.7591)$. The optimal bandwidths for the local linear approximation of $\widehat{U}_i^{\partial}$ by

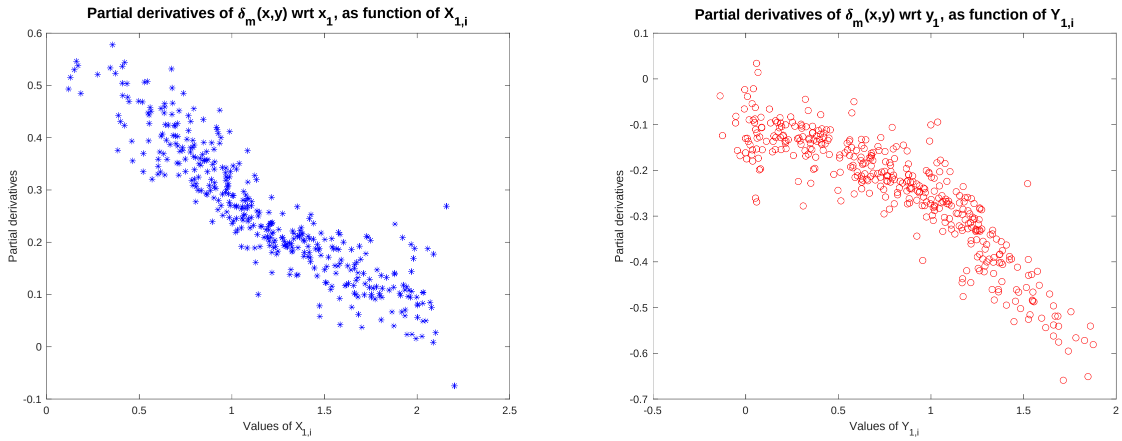

Figure 6: Nonlinear case: $\partial\delta_m(x, y)/\partial x_1$ and $\partial\delta_m(x, y)/\partial y_1$ as function of x_1 (left panel) and y_1 (right panel) respectively. Here $n = 400$, $m = 200$ with $p = q = 2$.

Table 6: Nonlinear example: bootstrap 95% confidence intervals of the partial derivatives $\partial\delta_m(w)/\partial w$ for the first 20 units in the sample. For each unit and each derivative we have the lower bound and the upper bound of the corresponding intervals. Here $n = 400$, $m = 200$ and $p = q = 2$.

	$\partial\delta_m(x,y)/\partial x_1$		$\partial\delta_m(x,y)/\partial x_2$		$\partial\delta_m(x,y)/\partial y_1$		$\partial\delta_m(x,y)/\partial y_2$	
1	0.2678	0.3163	0.3085	0.3685	-0.2336	-0.1954	-0.1798	-0.1267
2	0.2277	0.2688	0.2619	0.3035	-0.1843	-0.1556	-0.3237	-0.2772
3	0.0884	0.1495	0.0929	0.1771	-0.3671	-0.3112	-0.4437	-0.3713
4	0.1606	0.2380	0.1324	0.2063	-0.2072	-0.1383	-0.5011	-0.4102
5	0.4240	0.5190	0.1232	0.2676	-0.1937	-0.1131	-0.2490	-0.0949
6	0.4302	0.4858	0.2598	0.3122	-0.1513	-0.1149	-0.1464	-0.0969
7	0.1871	0.2459	0.2218	0.2707	-0.2479	-0.2038	-0.3371	-0.2833
8	0.1195	0.1641	0.3250	0.3983	-0.1405	-0.0976	-0.4116	-0.3379
9	0.1357	0.2907	0.2584	0.3432	-0.2418	-0.1584	-0.3442	-0.2347
10	0.2461	0.2959	0.1061	0.1629	-0.3098	-0.2558	-0.3494	-0.2847
11	0.3007	0.3936	0.1790	0.2497	-0.3439	-0.2367	-0.1993	-0.1063
12	0.0519	0.1511	0.1685	0.2446	-0.3827	-0.3273	-0.3831	-0.2823
13	0.0339	0.1292	0.1849	0.2409	-0.3807	-0.3178	-0.4035	-0.3185
14	0.2329	0.2816	0.3973	0.4893	-0.1189	-0.0692	-0.2560	-0.1500
15	-0.0210	0.0892	0.3258	0.4264	-0.4172	-0.3466	-0.2597	-0.1739
16	0.1878	0.2292	0.2965	0.3591	-0.2235	-0.1835	-0.2872	-0.2346
17	0.1756	0.2636	0.1476	0.2780	-0.1509	-0.0689	-0.5480	-0.3991
18	0.2511	0.2995	0.1280	0.1883	-0.1307	-0.0724	-0.4988	-0.4287
19	0.1362	0.2100	0.0753	0.1366	-0.2648	-0.2051	-0.5322	-0.4399
20	0.2482	0.2991	0.1087	0.1558	-0.4099	-0.3655	-0.2398	-0.1763

(V_i, Z_i) are given by (2.6416, 5.6873, 1.2370, 5.4950) and for the order-m frontier approximation of $\widehat{U}^{\partial}_{m,i}$ by (V_i, Z_i) we have (2.6444, 5.6683, 1.2427, 2.7212).

To illustrate the results we show some pictures of the obtained estimates and some tables with the estimated confidence intervals obtained by applying a basic bootstrap with $B = 1000$ replications.

Figure 7 shows the distribution of the frontier fits obtained by the local linear approximations for the FDH case (boxplot on the left) and for the order$-m$ case with $m = 550$ (boxplot on the right).

Figure 8 in the left panel, shows the distribution of the estimated marginal rates of substitution reporting the FDH case in boxplots 1 and 2 and the order$-m$ case with $m = 550$ in boxplots 3 and 4. Figure 8, right panel, illustrates the boxplots of the transformation rates estimated with FDH (boxplot 1) and order$-m$ with $m = 550$ (boxplot 2).

Figure 9 shows the boxplots of the estimated derivatives of the distance functions with respect to x, y_1 and y_2 estimated with FDH (boxplots 1, 2 and 3) and with order-m (boxplots 4, 5 and 6).

Figure 10 shows the plots of the estimated derivatives of the conditional distance functions $\delta_m(x, y|z)$ (i) w.r.t. x versus the observed values of x (left panel), (ii) w.r.t. y_1 versus the observed values of y_1 (middle panel) and (iii) w.r.t. y_2 versus the observed values of y_2 (right panel). Overall, we observe the expected signs for all the estimates illustrated in Figures 7: 10.

Table 8 shows the 95% bootstrap confidence intervals for the rates $\partial(y_1)/\partial x_1$ (second and third columns),

Table 7: Inputs, Outputs and Environmental variables used in the analysis: averages by country. Source: Daraio et al. (2021, p.8).

Country (Code)	#obs	ACAD	NONACAD	TEXP	TDEG	PUB	PHD	SPEC	SIZE
Belgium (BE)	5	297065158.52	2851.11	1425.54	5002.40	330.80	12685.60	0.50	9.40
Switzerland (CH)	11	334058152.35	2251.07	1115.27	2417.00	315.09	8884.00	0.65	9.23
Cyprus (CY)	1	107912583.34	389.00	566.00	1525.00	43.00	2862.00	0.71	8.81
Germany (DE)	73	431951227.71	2009.73	2458.53	2801.77	351.82	7352.86	0.65	9.63
Denmark (DK)	8	310038255.68	2106.25	1694.25	4023.50	193.12	8438.25	0.66	9.47
Hungary (HU)	7	300646738.90	1310.14	3225.29	4325.43	134.71	3927.14	0.70	10.05
Ireland (IE)	10	163345454.22	929.21	814.84	3826.90	136.80	4019.30	0.62	9.42
Italy (IT)	60	261088844.47	1448.43	959.42	4890.20	181.73	5989.38	0.67	9.87
Lithuania (LT)	4	354822075.78	955.25	927.75	3416.50	47.00	2292.00	0.80	9.41
Luxembourg (LU)	1	112655414.25	741.98	306.53	803.00	57.00	1643.00	0.74	8.46
Malta (MT)	1	104048589.04	791.00	759.00	3345.00	19.00	897.00	0.67	9.22
Netherlands (NL)	13	412966543.19	1868.71	1381.97	5836.38	295.46	14897.46	0.62	9.78
Norway (NO)	10	193484267.88	1298.14	938.31	2417.30	127.60	4955.40	0.70	8.94
Portugal (PT)	17	88363541.43	860.31	599.77	2757.24	115.06	2968.94	0.69	9.24
Sweden (SE)	20	224089812.76	1215.27	920.08	2478.10	166.30	6234.75	0.67	9.62
United Kingdom (UK)	96	279935092.91	1339.79	1517.76	6464.06	261.30	6801.55	0.64	9.73
Europe (EU)	337	298163446.67	1538.62	1506.55	4411.41	241.41	6760.74	0.66	9.64

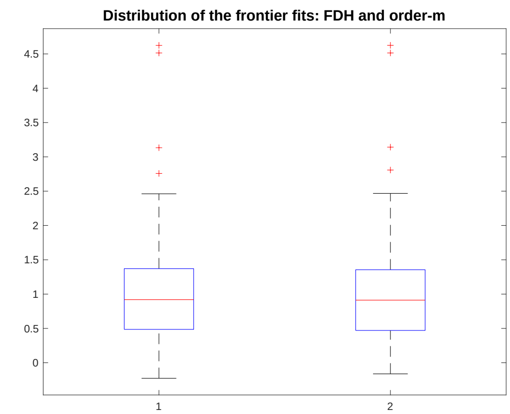

Figure 7: Higher Education Institutions (HEI) data: distribution of the frontier fits obtained by the local linear approximations. FDH case: boxplot 1 (on the left); order-m case with $m = 550$: boxplot 2 (on the right).

$\partial(y_2)/\partial x_1$ (fourth and fifth columns) and $\partial y_1/\partial y_2$ (sixth and seventh columns) evaluated at conditional order-m frontier points, with $m = 550$. The table shows the results of 20 observations to save space. We observe that for most observations the estimated results are statistically significant at 95% except for units #9, 15, 16, 18 and 19 which include zero in their confidence intervals.

Table 9 shows the results of the bootstrap confidence intervals for $\partial\delta_m(x, y|z)/\partial(x)$ (second and third columns), $\partial\delta_m(x, y|z)/\partial(y_1)$ (fourth and fifth columns) and $\partial\delta_m(x, y|z)/\partial(y_2)$ (sixth and seventh columns), with $m = 550$. The table, again, reports 20 institutions to save space. We observe that all institutions except unit # 19 (that includes zero in one of its estimated confidence intervals) show results statistically significant at 95%.

Overall, we observe that the estimated quantities showed in Tables 8 and 9 have the expected signs.

Note that the derivatives and the particular rates are in the units of the factors $X = FX$ and $Y_2 = FY$. In Appendix A we describe how these derivatives in "factor units" can provide the derivatives in the original units.

9 Conclusions

Nonparametric methods providing envelopment estimators, like the FDH or DEA are very attractive since they do not rely on restrictive parametric assumptions on the DGP, in particular the shape of the boundary

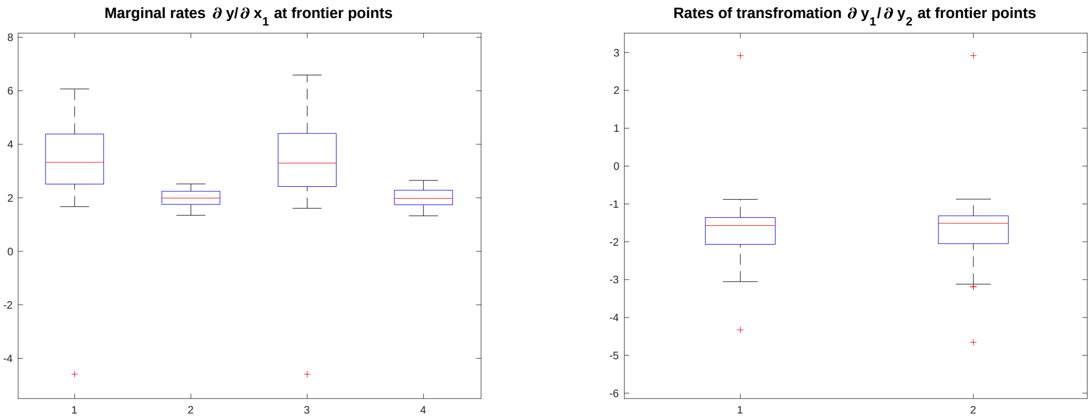

Figure 8: HEI data: distribution of the estimated marginal and transformation rates at frontier points. Left panel, marginal rates: FDH case $\partial y_1/\partial x$ (boxplot 1) and $\partial y_2/\partial x$ (boxplot 2), order-m case with $m = 550$: $\partial y_1/\partial x$ (boxplot 3) and $\partial y_2/\partial x$ (boxplot 4) . Right panel transformation rates $\partial y_1/\partial y_2$:FDH case (boxplot 1) and order-m case with $m = 550$ (boxplot 2).

and the distributions of inputs and outputs in the attainable set. However these nonparametric techniques do not allow to make sensitivity analysis of the estimated frontiers, e.g. derivatives of the optimal production outputs relative to some particular inputs, or to infer about marginal products and other coefficients of the frontier of economic interest. Parametric models for production frontiers permit easier economic interpretation but at a cost of restrictive assumptions on the DGP. In addition, these parametric models rely mostly on regression techniques fitting the center of the cloud of observed points instead of capturing its efficient boundary.

In this paper we propose an approach which avoids these drawbacks and provide approximations of the economic coefficients of interest by "smoothing" the nonparametric estimators of the frontiers. It is an extension of the ideas initiated by Florens and Simar (2005) who propose linear approximations of (univariate) frontier functions. The novelty of our approach is manifold. In particular, it allows to handle fully multivariate cases in a flexible underlying efficiency model based on directional distances and it provides also flexible approximations based on local linear tools providing local estimates of all the desired partial derivatives. Simple bootstrap algorithms allow for inference on all the coefficients of interest. We also extend the methods for dealing with environmental factors. Illustrations with some simulated data sets and with a real data from European universities show the usefulness and the flexibility of the proposed approach.

A Derivatives: from Factors to Original Units

Consider an output factor F_y defined as $F_y = a'\breve{y}$ for some $\breve{y} \in \mathbb{R}_+^L$ where $L \leq q$ is the number of outputs aggregated in F_y. We denote the eventual outputs in y not aggregated by $\mathring{y} \in \mathbb{R}_+^K$, where $K = q - L$, so $y = (\breve{y}' \; \mathring{y}')'$. We know that $a \in \mathbb{R}^L$, with $a'a = 1$ is the eigenvector of the 2nd moments matrix of the L outputs $\breve{y}$, corresponding to its largest eigenvalue.

Similarly we may have an input factor F_x defined as $F_x = b'\breve{x}$ for some $\breve{x} \in \mathbb{R}^J$ where $J \leq p$ is the number of inputs aggregated in F_x and $b \in \mathbb{R}^J$ is the eigenvector, with $b'b = 1$, of the second moment matrix of these J inputs $\breve{x}$, corresponding to its largest eigenvalue. Again we denote by $\mathring{x} \in \mathbb{R}_+^I$, where $I = p - J$, the inputs not aggregated by F_x so that $x = (\breve{x}' \; \mathring{x}')'$.

Suppose we have a procedure to evaluate (or to estimate) the derivatives of F_y w.r.t. some variables (another output in $\mathring{y}$ or some inputs in $\mathring{x}$ or to an input factor F_x) at the corresponding frontier point, say F_y. We know that this frontier point F_y has coordinates in the $\breve{y}$-space given by

$$\breve{y} = F_y a \in \mathbb{R}_+^L. \tag{A.1}$$

Consider now a differential ∂F_y relative to some other variable. This differential shifts the frontier point along the direction a at the point $F_y + \partial F_y$, which has coordinates in the $\breve{y}$-space given by $(F_y + \partial F_y)a = \breve{y} + \partial F_y a$,

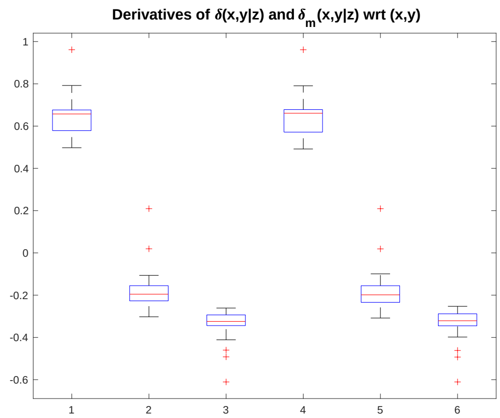

Figure 9: HEI data: distribution of the estimated derivatives of the distance functions with respect to x, y_1 and y_2. On the left, FDH case: estimated derivatives of $\delta(x, y|z)$ with respect to x (boxplot 1), w.r.t. y_1 (boxplot 2) and w.r.t. y_2 (boxplot 3). On the right, order-m case: estimated derivatives of $\delta_m(x, y|z)$ w.r.t. x (boxplot 4), w.r.t. y_1 (boxplot 5) and w.r.t. y_2 (boxplot 6).

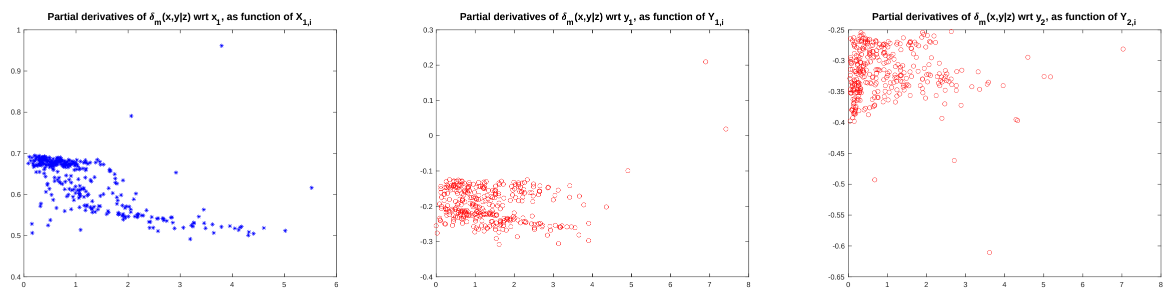

Figure 10: HEI data: plots of the estimated derivatives of $\delta_m(x, y|z)$, from left to right with respect to x, y_1 and y_2 versus the observed values of x, y_1 and y_2.

so that we may define the corresponding differential in the $\breve{y}$-space as

$$\partial \breve{y} = \partial F_y a \in \mathbb{R}^L. \tag{A.2}$$

We would obtain a similar result in the $\breve{x}$-space when considering a differential of F_x relative to some variables and define

$$\partial \breve{x} = \partial F_x b \in \mathbb{R}^J. \tag{A.3}$$

Now it is easy to consider various derivatives involving the factors F_y or F_x or both. For instance, by simple algebra we may have the derivatives of the component of an output factor F_y relative to $\mathring{x}$

$$\frac{\partial \breve{y}}{\partial \mathring{x}'} = \frac{\partial F_y a}{\partial \mathring{x}'} = a \frac{\partial F_y}{\partial \mathring{x}'}, \tag{A.4}$$

i.e. a $(L \times I)$ matrix with element (ℓ, i) given by $a_\ell(\partial F_y / \partial \mathring{x}_i)$.

Another case is to consider the derivatives of $\mathring{y}$ with respect to the components of an input factor F_x. Here we have

$$\frac{\partial \mathring{y}}{\partial \breve{x}'} = \frac{\partial \mathring{y}}{\partial F_x b'} = \frac{\partial \mathring{y}}{\partial F_x}(i_J \oslash b)', \tag{A.5}$$

where i_L is a L-vector of ones and $\oslash$ stands for the Hadamard division of vectors (element-wise). So we have a $(K \times J)$ matrix with (k, j) element $(\partial \mathring{y}_k / \partial F_x)(1/b_j)$.

Table 8: HEI data: 95% bootstrap confidence intervals (lower and upper bounds) for the rates $\partial(y_1)/\partial x_1$ (second and third columns), $\partial(y_2)/\partial x_1$ (fourth and fifth columns) and $\partial y_1/\partial y_2$ (sixth and seventh columns) evaluated at conditional order-m frontier points, with $m = 550$.

	$\partial y_1/\partial x_1$		$\partial y_2/\partial x_1$		$\partial y_1/\partial y_2$	
1	0.4923	2.9147	1.2657	2.3184	-1.9174	1.3169
2	2.9255	4.2403	2.1341	3.2376	-1.7754	-0.6607
3	2.2144	3.3965	1.8728	2.4902	-1.6576	-0.5779
4	1.2743	2.8987	1.4926	2.3841	-1.7618	0.2435
5	1.7150	3.3900	1.7738	2.3769	-1.7260	-0.1620
6	1.5881	3.3130	1.6131	2.2146	-1.8981	-0.1620
7	1.5249	3.2928	1.5517	2.1992	-1.9720	-0.1238
8	2.7499	3.6280	2.1313	3.0884	-1.5334	-0.5831
9	1.3585	3.1614	1.5194	2.2331	-1.9087	0.1000
10	1.7500	3.3344	1.6385	2.2320	-1.8971	-0.3550
11	3.8307	4.9358	2.2365	3.1665	-2.0822	-0.7821
12	2.9885	4.2796	2.1723	3.6854	-1.7294	-0.4020
13	1.0276	2.9885	1.5033	2.4155	-1.7418	0.4673
14	3.7620	4.8902	2.2816	3.2678	-2.0211	-0.6914
15	1.2028	3.2295	1.5200	2.1238	-1.9903	0.0808
16	-1.4520	2.8425	1.4608	2.5791	-1.6573	3.8764
17	2.8942	3.8955	2.2305	3.3608	-1.5809	-0.5416
18	-0.7522	2.9323	1.2525	2.4006	-1.8790	3.0719
19	-5.4514	3.0516	1.0189	2.4379	-2.0092	8.3581
20	2.7563	3.8669	2.1203	3.3175	-1.6175	-0.5227

Table 9: HEI data: 95% bootstrap confidence intervals (lower and upper bounds) for $\partial\delta_m(x, y|z)/\partial(x)$ (second and third columns), $\partial\delta_m(x, y|z)/\partial(y_1)$ (fourth and fifth columns) and $\partial\delta_m(x, y|z)/\partial(y_2)$ (sixth and seventh columns), with $m = 550$.

	$\partial\delta_m(x, y\|z)/\partial x$		$\partial\delta_m(x, y\|z)/\partial y_1$		$\partial\delta_m(x, y\|z)/\partial y_2$	
1	0.4923	0.6628	-0.3707	-0.1481	-0.4237	-0.0722
2	0.6472	0.7515	-0.2232	-0.1481	-0.3102	-0.1654
3	0.5917	0.6776	-0.2549	-0.1843	-0.3264	-0.2127
4	0.5259	0.6477	-0.3293	-0.1715	-0.3842	-0.1324
5	0.5786	0.6682	-0.2770	-0.1825	-0.3382	-0.2041
6	0.5574	0.6580	-0.2879	-0.1677	-0.3697	-0.2118
7	0.5461	0.6641	-0.2962	-0.1537	-0.3854	-0.2040
8	0.6289	0.7139	-0.2389	-0.1784	-0.3008	-0.1690
9	0.5348	0.6596	-0.3114	-0.1606	-0.3883	-0.1821
10	0.5628	0.6604	-0.2782	-0.1674	-0.3662	-0.2206
11	0.6898	0.7621	-0.1847	-0.1412	-0.3120	-0.1842
12	0.6603	0.7680	-0.2302	-0.1396	-0.3074	-0.1271
13	0.5234	0.6609	-0.3453	-0.1623	-0.3802	-0.1338
14	0.6908	0.7698	-0.1876	-0.1425	-0.3065	-0.1721
15	0.5313	0.6583	-0.3119	-0.1505	-0.3924	-0.1946
16	0.4777	0.6696	-0.4508	-0.1621	-0.3830	0.0055
17	0.6453	0.7375	-0.2328	-0.1660	-0.2949	-0.1499
18	0.4713	0.6738	-0.4217	-0.1460	-0.4243	-0.0165
19	0.4264	0.7143	-0.4859	-0.1229	-0.4437	0.0769
20	0.6370	0.7376	-0.2389	-0.1579	-0.3054	-0.1491

If we want to recover the derivatives of $\check{y}$ with respect to the elements of $\check{x}$, they are given by

$$\frac{\partial \check{y}}{\partial \check{x}'} = \frac{\partial F_y a}{\partial F_x b'} = \frac{\partial F_y}{\partial F_x} a (i_J \oslash b)'. \tag{A.6}$$

This is a $(L \times J)$ matrix with (ℓ, j) element $(\partial F_y/\partial F_x)a_\ell/b_j$.

Note that we have another useful consequence. The elasticities in terms of the factors are recovered in terms of their components. This is due to the fact that $\partial\check{y} \oslash \check{y} = (a\partial F_y) \oslash (aF_y) = \partial F_y/F_y$. This is because we distribute the differential ∂F_y in the $\check{y}$-space in the direction of a, i.e. in the direction of $\check{y}$, i.e. in a radial proportional way. So it is easy to check that for instance

$$\mathcal{E}(\check{y}_\ell, \check{x}_j) = \mathcal{E}(F_y, F_x), \tag{A.7}$$

or, for another instance

$$\mathcal{E}(\check{y}_\ell, \mathring{x}_i) = \mathcal{E}(F_y, \mathring{x}_i), \tag{A.8}$$

and many other possibilities.

We are also able to find the marginal rate of substitutions on the frontier, between inputs in $\breve{x}$ and inputs in $\mathring{x}$ or the rate of transformation on the frontier between outputs in $\breve{y}$ and inputs in $\mathring{y}$. For instance

$$\frac{\partial \breve{y}}{\partial \mathring{y}'} = \frac{\partial F_y a}{\partial \mathring{y}'} = a\frac{\partial F_y}{\partial \mathring{y}'}, \tag{A.9}$$

i.e. a $(L \times K)$ matrix with element (ℓ, k) given by $a_\ell(\partial F_y/\partial \mathring{y}_k)$.

Note that this approach does not allow to recover from the factors the marginal rates of substitution between the elements $\breve{x}$ composing the factor F_x, or the marginal rates of transformation between the outputs $\breve{y}$ composing the factor F_y. For instance the information about $\partial \breve{y}_{\ell_1}/\partial \breve{y}_{\ell_2}$ on the frontier is lost if we only have the factor F_y. Using the ideas above would provide the trivial value a_{ℓ_1}/a_{ℓ_2}, which corresponds to the radial ratio considered above and has no economic interest.

References

[1] Afriat, S.N. (1972), Efficiency estimation of production functions, *International Economic Review*, 13, 568–598.

[2] Aragon, Y. and Daouia, A. and Thomas-Agnan, C. (2005), Nonparametric Frontier Estimation: A Conditional Quantile-based Approach, *Econometric Theory*, 21, 358–389.

[3] Baker, B. D. (2001). Can flexible non-linear modeling tell us anything new about educational productivity? *Economics of Education Review*, 20(1), 81-92.

[4] Bonaccorsi, A., Daraio, C. (Eds.). (2007). *Universities and strategic knowledge creation: Specialization and performance in Europe*. Edward Elgar Publishing.

[5] Bonaccorsi A., Daraio C., Simar, L. (2007), Efficiency and productivity in European Universities. Exploring trade-offs in the strategic profile, in Bonaccorsi A. and Daraio C. (Eds.) *Universities and strategic knowledge creation: Specialization and performance in Europe*. Edward Elgar Publishing., pp. 144-208.

[6] Cazals, C. Florens, J.P. and L. Simar (2002), Nonparametric Frontier Estimation: a Robust Approach , in *Journal of Econometrics*, 106, 1–25.

[7] Chambers, R. G., Y. Chung, and R. Färe (1998), Profit, directional distance functions, and nerlovian efficiency, *Journal of Optimization Theory and Applications*, 98, 351–364.

[8] Coelli, T. and S. Perelman (1999), A comparison of parametric and non-parametric distance functions: With application to European railways, *European Journal of Operational Research*, 117, 326–339.

[9] Daouia, A., J.P. Florens and L. Simar (2008), Functional Convergence of Quantile-type Frontiers with Application to Parametric Approximations, *Journal of Statistical Planning and Inference*, 138, 708–725.

[10] Daouia, A., J.P. Florens and L. Simar (2012), Regularization of nonparametric frontiers estimators, *Journal of Econometrics*, 168, 285–299.

[11] Daouia, A. and L. Simar (2007), Nonparametric efficiency analysis: a multivariate conditional quantile approach, *Journal of Econometrics*, 140, 375–400.

[12] Daraio, C. and L. Simar (2005), Introducing environmental variables in nonparametric frontier models: a probabilistic approach, *Journal of Productivity Analysis*, vol 24, 1, 93–121.

[13] Daraio, C. and L. Simar (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*, Springer, New-York.

[14] Daraio, C., Simar, L. and P.W. Wilson (2018), Central Limit Theorems for Conditional Efficiency Measures and Tests of the "Separability" Condition in Nonparametric, Two-Stage Models of Production, *Econometrics Journal*, 21, 170–191.

[15] Daraio, C., Simar, L. and P.W. Wilson (2020), Fast and Efficient Computations of Directional Distances Estimators. *Annals of Operations Research*, 288, 805–835.

[16] Daraio, C., Simar, L., Wilson, P. W. (2021). Quality as a latent heterogeneity factor in the efficiency of universities. *Economic Modelling*, 99, 105485.

[17] Debreu, G. (1951), The coefficient of resource utilization, *Econometrica*, 19:3, 273-292.

[18] De Witte, K., Lopez-Torres, L. (2017). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society*, 68(4), 339-363.

[19] Dewey, J., Husted, T. A., Kenny, L. W. (2000). The ineffectiveness of school inputs: a product of misspecification?. *Economics of Education Review*, 19(1), 27-45.

[20] Färe, R., Grosskopf, S. and C.A.K. Lovell (1985), *The Measurement of Efficiency of Production.* Boston, Kluwer-Nijhoff Publishing.

[21] Fan, J. and I. Gijbels, (1996), *Local Polynomial Modelling and Its Applications*, Chapman & Hall, London.

[22] Farrell, M.J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society*, A(120), 253-281.

[23] Figlio, D. N. (1999). Functional form and the estimated effects of school resources. *Economics of education review*, 18(2), 241-252.

[24] Florens, J.P. and L. Simar, (2005), Parametric Approximations of Nonparametric Frontier, *Journal of Econometrics*, vol 124, 1, 91–116

[25] Greene, W.H. (1990), A gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46, 141–163.

[26] Grosskopf, S., Hayes, K., Taylor, L. L. (2014). Applied efficiency analysis in education. *Economics and Business Letters*, 3(1), 19-26

[27] Grosskopf, S., Hayes, K., Taylor, L. and W. Weber (1997), Budget constrained frontier measures of fiscal equality and efficiency in schooling, *Review of Economics and Statistics*, 79, 116–124.

[28] Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of human Resources*, 351–388.

[29] Hanushek, E.A. (1986) The economics of schooling, production and efficiency in public schools, *Journal of Economic Literature*, 24, 11411177.

[30] Härdle, W. and L. Simar (2019), *Applied Multivariate Statistical Analysis*, Fifth Edition, 558p., Springer-Verlag, Berlin.

[31] Jeong, S.O. , B. U. Park and L. Simar (2010), Nonparametric conditional efficiency measures: asymptotic properties. *Annals of Operations Research*, 173, 105–122.

[32] Jeong, S.O. and L. Simar (2006), Linearly interpolated FDH efficiency score for nonconvex frontiers, *Journal of Multivariate Analysis*, 97, 2141–2161.

[33] Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, 25(3), 273–288.

[34] Kneip, A, L. Simar and P. W. Wilson (2008), Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models, *Econometric Theory*, 24, 1663–1697.

[35] Kneip, A., Simar, L. and P. W. Wilson (2015), When bias kills the variance: Central Limit Theorems for DEA and FDH efficiency scores, *Econometric Theory*, 31, 394–422.

[36] 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 Commision for Research in Economics, Monograph 13, John-Wiley, New-York.

[37] Kumbhakar, S.C. and C.A.K. Lovell (2000), *Stochastic Frontier Analysis*, Cambridge University Press.

[38] Park, B. Simar, L. and Ch. Weiner (2000), The FDH Estimator for Productivity Efficiency Scores: Asymptotic Properties, *Econometric Theory*, Vol 16, 855–877.

[39] Ruggiero, J. (2004). Performance evaluation in education. *Handbook on data envelopment analysis*, 323–346.

[40] Shephard, R.W. (1970). *Theory of Cost and Production Function.* Princeton University Press, Princeton, New-Jersey.

[41] Simar, L. and A. Vanhems (2012), Probabilistic Characterization of Directional Distances and their Robust versions, *Journal of Econometrics*, 166, 342–354.

[42] Simar, L., Vanhems, A. and P. W. Wilson (2012), Statistical inference with DEA estimators of directional distances, *European Journal of Operational Research*, 220, 853–864.

[43] Simar, L., Van Keilegom, I. and V. Zelenyuk (2017), Nonparametric Least Squares Methods for Stochastic Frontier Models, *Journal of Productivity Analysis*, 47, 189–204.

[44] Simar, L and P. Wilson (2007), Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes, *Journal of Econometrics*, vol 136, 1, 31–64.

[45] Simar, L. and P.W. Wilson (2008), Statistical Inference in Nonparametric Frontier Models: recent developments and perspectives, in *The Measurement of Productive Efficiency and Productivity Growth*, 2nd Edition, Harold Fried, C.A.Knox Lovell and Shelton Schmidt, editors, Oxford University Press, 421–521.

[46] Simar, L. and P. W. Wilson (2013), Estimation and inference in nonparametric frontier models: Recent developments and perspectives, *Foundations and Trends in Econometrics*, Vol. 5: No 3-4, pp 183–337.

[47] Simar, L. and P. W. Wilson (2015), Statistical Approaches for Nonparametric Frontier Models: A Guided Tour, *International Statistical Review*, 83, 1, 77–110.

[48] Simar, L. and P. W. Wilson (2020), Hypothesis Testing in Nonparametric Models of Production using Multiple Sample Splits. *Journal of Productivity Analysis*, 53, 287–303.

[49] Simar, L. and P. W. Wilson (2021), Nonparametric, Stochastic Frontier Models with Multiple Inputs and Outputs, LIDAM Discussion paper ISBA 2021/03, UCL.

[50] Stevenson, R.E. (1980), Likelihood functions for generalized stochastic frontier estimation, *Journal of Econometrics*, 13, 57–66.

[51] Stock, J. H. (2010). The other transformation in econometric practice: Robust tools for inference. *The Journal of Economic Perspectives*, 24(2), 83–94.

[52] Wilson, P.W. (2008), FEAR: A Software Package for Frontier Efficiency Analysis with R, *Socio-Economic Planning Sciences* 42, 247–254.

[53] Wilson, P. W. (2018), Dimension reduction in nonparametric models of production, *European Journal of Operational Research*, 267, 349–367.