Journal of Productivity Analysis (2024) 62:197–215
https://doi.org/10.1007/s11123-024-00726-2

Approximations and inference for envelopment estimators of production frontiers

Cinzia Daraio[1] · Léopold Simar[2]

Accepted: 21 May 2024 / Published online: 2 July 2024

Abstract

Nonparametric methods have been commonly used to assess the performance of both private and public organizations. Among them, the most popular ones are envelopment estimators such as Free Disposal Hull (FDH) or Data Envelopment Analysis (DEA), which estimate the attainable sets and their efficient boundaries by enveloping the cloud of observed units in the appropriate input-output space. However, these nonparametric envelopment techniques do not provide estimates of marginal products and other coefficients of economic interest. This paper presents a new approach that provides local estimates of all the desired partial derivatives and economic coefficients, which complement and complete the analysis based on nonparametric envelopment estimators. We improve nonparametric estimators by estimating nonparametrically *smoothed* efficient boundaries and providing derivatives and other coefficients without having to assume any parametric structure for the frontier and the inefficiency distribution. Our approach offers several advantages, such as a flexible nonparametric adjustment of the efficient frontier based on local linear models; a general multivariate efficiency model based on directional distances where one can choose the desired benchmark direction; the possibility of assessing the impact of external-environmental variables; a bootstrap-based statistical inference for deriving confidence intervals on the estimated coefficients for nonparametric and robust frontier approximations; the possibility of including factors aggregating inputs or outputs and recovering the estimated coefficients in the original units. To demonstrate the usefulness of the proposed approach, we provide an illustration in the field of education, where economic coefficients are important but the parametric assumptions have been questioned.

Keywords Data envelopment analysis · Partial frontiers · Directional distances · Linear approximations · Local linear approximations

JEL classification C1 · C14 · C13

1 Introduction and contribution

Efficiency analysis examines how production units transform their inputs into outputs, that may be goods or services. Nonparametric envelopment estimators are highly preferred as they require very few assumptions. These estimators do not require any particular shape for the attainable set and its frontier, except for free disposability[1] for the Free Disposal Hull (FDH) and free disposability and convexity of the attainable set for Data Envelopment Analysis (DEA). Additionally, they do not require any specific distributional assumption for the distribution of inefficiencies. The statistical properties of these envelopment estimators have been established and inference is available (see Simar and Wilson 2013, 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(s) to particular inputs, marginal rates of substitution between inputs, marginal rates of transformation between outputs and so on.

✉ Cinzia Daraio
daraio@diag.uniroma1.it

[1] Department of Computer, Control and Management Engineering A. Ruberti (DIAG), Sapienza University of Rome, via Ariosto 25, 00185 Rome, Italy

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

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

In existing literature, the parametric and nonparametric approaches to efficiency analysis compete and are applied separately, each with its own advantages and disadvantages. The parametric approach offers the possibility of estimating economic coefficients and substitution rates at the cost of imposing specific functional forms for the frontier and for the inefficiency distributions. The nonparametric approach, on the other hand, is more general because it does not impose assumptions on the functional forms of the frontier and the inefficiency distribution, but at the cost of not providing the user with useful quantitative information such as relevant coefficients and substitution rates. The method we propose in this paper allows the nonparametric approach to be reinforced by giving the possibility of estimating economic coefficients by flexible approximations of non-parametric envelopment frontiers.

We introduce now the basic concepts and notations. The efficient production frontier is defined in the appropriate input-output space as the locus of the optimal combination 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 Ψ embodies the fundamental features that are commonly found in economic theory, as described in Shephard (1970). The efficient boundary (frontier) of this set is the set of efficient combinations of inputs and outputs

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

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 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 "Oriented radial measures"). 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, order-m frontiers can also be considered. This allows to define $\delta_m(x, y)$ which by construction are smaller than $\delta(x, y)$, unless $m \to \infty$ (see Simar and Vanhems 2012). These efficiency measures benchmark a unit (x, y) against less extreme frontiers, and so share robustness properties, robustness against outliers or extreme observations. The "partial frontier" of order-m points are defined as

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

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 popular nonparametric estimators of Ψ are based on envelopment estimators, like FDH or DEA estimators. From these estimators, it is easy to define for any (x, y), $\widehat{\delta}_n(x, y)$, a nonparametric estimator of $\delta(x, y)$. Estimators of partial efficiency scores $\widehat{\delta}_{m,n}(x, y)$ 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]

Florens and Simar (2005) (hereafter FS) have suggested to approximate nonparametric estimators of the frontier by a linear model, avoiding specification of a parametric family of densities for the stochastic part of the model and fitting the cloud of points near the efficient boundary. However, FS considered the univariate case, analyzed only FDH and order-m frontiers in the first-stage and proposed only linear models for the approximation in the second-stage.

The difficulty of the multivariate setup, where $r = p + q$, is that the efficient frontier, in the (x, y)-space, is $(r - 1)$-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 needed 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 build estimators of Stochastic Frontier Analysis (SFA) in a multivariate setup in Simar and Wilson (2022) (hereafter SW). SFA relies on a different paradigm than the one used in our paper, since it allows the presence of noise. Due to that, SFA mainly belongs to the family of parametric approaches in efficiency analysis, within which it is necessary to specify a functional form for the frontier, a functional form for the distribution of inefficiency, a functional form for the noise and some

[2] Daraio et al. (2020) provide the Matlab codes for the needed computations.

relationship between inefficiency and noise to achieve identification of the models. And even in this case, some intrinsic issues remain for estimating these models in particular due to the skewness of the OLS residuals (see e.g. Hafner et al. 2018).

Semiparametric approaches to SFA have been investigated e.g. by Simar et al. (2017) and extended to multivariate cases by SW, but there the Data Generating Process (DGP) requires also local parametric specification of the efficiency distribution and noise, to allow identification (see SW for a detailed discussion and further references). Within another semiparametric framework, Kuosmanen and Johnson (2017) introduced directional distances in a stochastic nonparametric envelopment framework, but adding some assumptions for the efficiency distribution and the noise, imposing convexity and shape restrictions, and finally providing only aggregate (average) coefficients.

In the traditional nonparametric envelopment approach (i.e. the so-called “deterministic” frontier models, to which DEA and FDH belong), the paradigm is different. Within this nonparametric envelopment approach, Podinovski (2019) extended a linear programming approach, applicable to any polyhedral production technology, incorporating undesirable outputs to estimate marginal characteristics of nonparametric production frontiers, including various marginal rates and elasticity measures. However, this approach assumes convexity, is not implemented for robust and nonconvex nonparametric efficiency measures, and does not provide any inference for the estimated coefficient of economic interest. Contrary to the existing literature, we provide a flexible nonparametric approximation of the traditional nonparametric envelopment estimators of the frontier and their robust and nonconvex versions, from which we can recover the coefficients and partial derivatives of economic interest, providing confidence intervals on the estimated coefficients and introducing external environmental variables in the analysis.

In this paper, we use a different DGP compared to semiparametric frameworks, which avoids identification and estimation issues. We will then adapt the DGP described in SW to our setup of nonparametric efficiency analysis, and we will derive new explicit relations between the characteristics in both coordinate systems. Then we will show how the strategy of FS can be extended to our more general setup. This will allow us to capture the shape of the cloud of points near its efficient boundary, without specifying any functional form. The main aim of this paper is to propose a two-stage approach that combines the flexibility of a first-stage based on nonparametric envelopment models (like FDH, DEA and partial efficiency measures) with a flexible second-stage based on a *smoothed* adjustment of the nonparametric efficient frontier estimated in the first-stage.

The idea is to “smooth” by some appropriate models (i.e., local linear models) the usual nonparametric estimators of the frontier without assuming any arbitrary frontier function. We will present our approach in a directional distance setup, and we will show how to adapt the approach to any other measure of efficiency (hyperbolic, input or output radial oriented). We will show that flexible local linear approximations are easy and feasible to handle, providing approximations of all the coefficients of economic interest, including derivatives, despite their nonparametric nature. We will provide guidelines for statistical inference for both the full and the order-m frontiers[3] approximations in order to derive confidence intervals on the estimated coefficients of interest. Finally, we will illustrate how the approach can easily be extended to deal with environmental factors.

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 computer power and 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. In this paper, we aim to build on the work of Stock (2010) by suggesting a method that combines nonparametric and robust approaches which include directional distances, with nonparametric local linear models. This approach will help us obtain more reliable coefficients and economic measurements that are not dependent on arbitrary functional forms.

The paper is organized as follows. Section “The statistical model and the transformation” illustrates the underlying statistical model and the transformation necessary to implement the fully multi-input multi-output case. Section “Our methodology” presents our methodology: (i) the best approximation and its estimation in Section “Best approximation and estimation”; (ii) the local linear approximation in Section “Local linear approximation” (iii) how to derive coefficients from factors to original units in Section “Derivatives: from factors to original units”; (iv) how to extend the approach when environmental factors may influence the frontier in Section “Dealing with environmental factors”, and the boostrap-based practical inference in Section “Bootstrap-based practical inference”. Section “Application on European universities” reports an illustration on real data, and Section ”Conclusions” concludes the paper. Supplementary Materials (SM) report introductory

[3] We focus the presentation to partial frontiers of order-m to save space. The extension, *mutatis mutandis*, to order-α frontiers is immediate and left to the readers.

information, simulated examples and additional information for interested readers.

2 The statistical model and the transformation

The definition of a clear DGP before introducing the estimation issues is crucial to understand what is the model we are analyzing. Moreover, in this section we describe how the rotation method of SW should be adapted to our setup, which is different from the one in SW (based on a different DGP). In particular, we give explicit equations for the links between the 2 spaces, which was not provided by SW. This is done in Eq. (2.16) which is new and useful for the developments of the methodology proposed in this paper.

2.1 The DGP and the rotation

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,

$$\begin{bmatrix} X_i \\ Y_i \end{bmatrix} = \begin{bmatrix} X_i^{\partial} \\ Y_i^{\partial} \end{bmatrix} - \delta_i \begin{bmatrix} -d_x \\ d_y \end{bmatrix}, \tag{2.1}$$

where the δ_i are conditionally to W_i^{∂} independent with $\delta_i | W_i^{\partial} \sim D_+\big(\eta(W_i^{\partial})\big)$ 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, including order-m. 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 $n \to \infty$,

$$n^{\kappa}\Big(\widehat{\delta}_{\bullet}(x, y) - \delta(x, y)\Big) \xrightarrow{\mathcal{L}} F_{\bullet}(\xi_{x,y}), \tag{2.2}$$

where κ determines the rate of convergence and depends on the assumptions on Ψ and the chosen estimator with "•" 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 (CLTs) 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 lose 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 order-m 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 $n \to \infty$,

$$\sqrt{n}\Big(\widehat{\delta}_m(x, y) - \delta_m(x, y)\Big) \xrightarrow{\mathcal{L}} N(0, \sigma^2_{m,x,y}), \tag{2.3}$$

where $\sigma^2_{m,x,y}$ has a known expression depending on the characteristics of the DGP. So inference with these estimators is much more easy. We will come back to all these nonparametric estimators below.

A natural way to transform the coordinates $w = (x, y)$ of the original space into a new system where the distance to the frontier (defined in (2.1)) 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 SW this is achieved by considering an arbitrary, but fixed, orthonormal basis for the direction vector d.[4] Let $S_d = [s_1 \ldots s_{r-1}]$ 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 $j = 1, \ldots, r-1$.

[4] 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 statistical packages have a build-in function to obtain S_d. In any case, an easy to program algorithm is described in Jeong and Simar (2006).

Clearly $S_d' S_d = I_{r-1}$ and $S_d' d = 0_{r-1}$.[5] Now we can define the $r \times r$ rotation matrix

$$R_d = \begin{bmatrix} S_d' \\ d'/||d|| \end{bmatrix}, \text{ and its transpose } R_d' = [S_d \quad d/||d||], \tag{2.4}$$

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

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

$$g_d : w \mapsto t = R_d w, \tag{2.5}$$

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 $r-1$ 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

$$\Gamma_d = \{t \in \mathbb{R}^r | t = g_d(w), w \in \Psi\}. \tag{2.6}$$

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

$$\phi(v) = \sup\{u | t = (v' \ u)' \in \Gamma_d\}, \tag{2.7}$$

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

$$\Gamma_d = \{t = (v' \ u)' \in \mathbb{R}^r | u \le \phi(v)\}. \tag{2.8}$$

For illustration, Fig. 1 shows this rotation for a simple case $p = q = 1$. Figure 1 displays a particular simulated dataset and the frontier points both in the original (x, y) space (left panel) and in the transformed (v, u) space (right panel). In the right panel of Fig. 1, u is in the direction of d and v is orthogonal to u.

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

$$\begin{bmatrix} V_i \\ U_i \end{bmatrix} = R_d W_i = \begin{bmatrix} S_d' W_i \\ ||d||^{-1} d' W_i \end{bmatrix}, \tag{2.9}$$

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

and the inverse relation between the observations is

$$W_i = R_d' \begin{bmatrix} V_i \\ U_i \end{bmatrix} = \left[S_d V_i + ||d||^{-1} d \, U_i \right]. \tag{2.10}$$

Hence our model (2.1) is transformed into

$$\begin{bmatrix} V_i \\ U_i \end{bmatrix} = \begin{bmatrix} V_i^\partial \\ U_i^\partial \end{bmatrix} - \delta_i \begin{bmatrix} 0_{r-1} \\ ||d|| \end{bmatrix}, , \tag{2.11}$$

or component by component

$$\begin{matrix} V_i = V_i^\partial \\ U_i = U_i^\partial - ||d||\delta_i \end{matrix}, \tag{2.12}$$

or simply due to (2.7)

$$U_i = \phi(V_i) - ||d||\delta_i, \tag{2.13}$$

where the heteroskedastic nature of δ in (2.1) can now be expressed in terms of V_i, since $W_i^\partial = R_d'[V_i' \phi(V_i)]'$. So (2.13) provides the scalar-valued equation to be estimated.

2.2 Some relations between the two spaces

We will see in Section “Best approximation and estimation” 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.

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

$$w^\partial = \begin{bmatrix} x^\partial \\ y^\partial \end{bmatrix} = [S_d \quad d/||d||] \begin{bmatrix} v \\ \phi(v) \end{bmatrix}, \tag{2.14}$$

i.e. a mapping from $\mathbb{R}^{r-1}$ to $\mathbb{R}^r$. One of the interests 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 $t_0^\partial = (v_0' \ u_0^\partial)'$ with $u_0^\partial = \phi(v_0)$ is given by the equation

$$c_t'(t - t_0^\partial) = 0, \text{ where } c_t' = [\nabla' \phi(v_0) \quad -1], \tag{2.15}$$

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

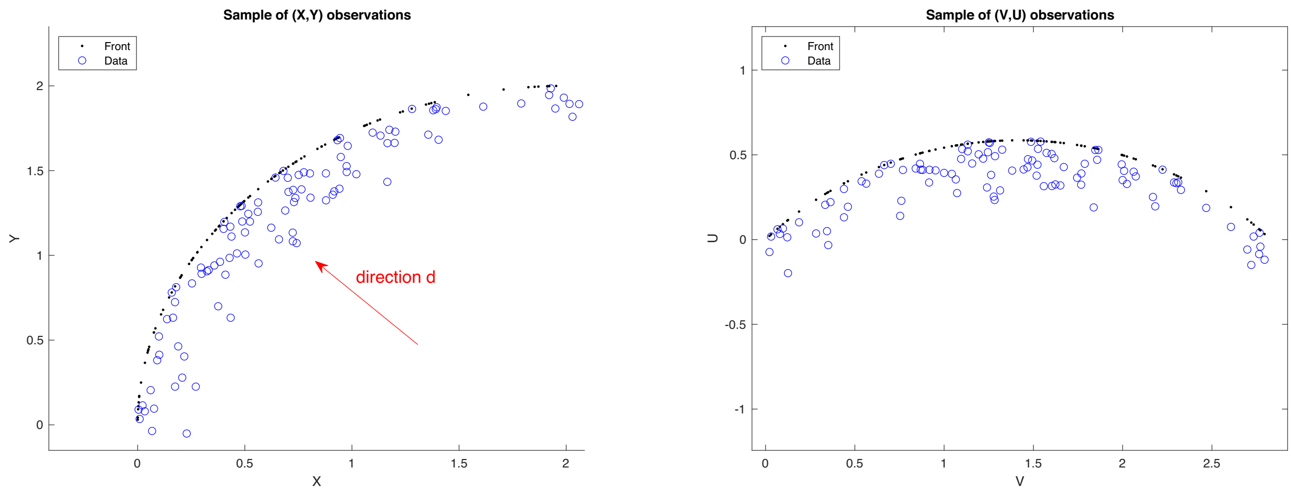

Fig. 1 An illustration of rotation in a simulated data set of size $n = 100$ with $p = q = 1$

hyperplane has equation $c_t' R_d(w - w_0^\partial)$, i.e.

$$c_w'(w - w_0^\partial) = 0, \text{ where } c_w = R_d' \begin{bmatrix} \nabla \phi(v_0) \\ -1 \end{bmatrix} \text{ and } v_0 = S_d' w_0^\partial. \tag{2.16}$$

Clearly the vectors c_t and c_w are given at a multiplicative constant ($\neq 0$).[6]

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

$$\left.\frac{\partial w_\ell}{\partial w_k}\right|_{w=w_0^\partial} = -\frac{c_k}{c_\ell}, \tag{2.17}$$

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 (2.17) is not defined, however, this indicates that w_ℓ has no effect on the frontier at this particular frontier point. So, if $c_k \neq 0$

$$\left.\frac{\partial w_k}{\partial w_\ell}\right|_{w=w_0^\partial} = 0. \tag{2.18}$$

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 (2.13)

$$\delta(v, u) = ||d||^{-1}(\phi(v) - u), \tag{2.19}$$

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

$$\delta(w) = ||d||^{-1}\left(\phi(S_d' w) - d'w/||d||\right). \tag{2.20}$$

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 (2.14) we have $w^\partial = S_d v + d\phi(v)/||d||$ then by plugging in $v = S_d' w$ we obtain (2.20).[7]

Note that we can also obtain the partial derivatives

$$\frac{\partial \delta(w)}{\partial w} = ||d||^{-1}\left(S_d \frac{\partial \phi(v)}{\partial v'} - \frac{d}{||d||}\right), \tag{2.21}$$

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 (2.14) can also be given by

$$\phi(S_d' w^\partial) - d' w^\partial / ||d|| = 0, \tag{2.22}$$

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

Interested readers can find two simple examples in Section A of the SM of this paper.

2.3 Oriented radial measures

To be exhaustive we summarize here the point of SW (see Section 3.5.1 in SW) showing that the directional model

[6] Note that we provide by (2.16) 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 (2.14), which was the way chosen by SW to characterize implicitly the vector c_w.

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

(2.1) can be used for handling radial distance functions. Suppose we are interested in the hyperbolic measures of efficiency (see Färe et al. 1985)

$$\tau(x,y)=\inf\{\tau|(\tau x,\tau^{-1}y)\in\Psi\}. \tag{2.23}$$

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 (2.1)

$$\begin{bmatrix}\log X_i\\ \log Y_i\end{bmatrix}=\begin{bmatrix}\log X_i^{\partial}\\ \log Y_i^{\partial}\end{bmatrix}-\delta_i\begin{bmatrix}-i_p\\ i_q\end{bmatrix}, \tag{2.24}$$

where i_k is a k-vector of ones. Then we obtain $\tau_i=\exp(\delta_i)$. Hence, we are back to our model (2.1), in the log-scale and using the direction vector as in (2.24).

As shown in SW, similar transformations allow us to consider the radial input efficiency (setting $d=[-i_p'\,0_q']'$ in (2.24)) and the radial output efficiency (setting $d=[0_p'\,i_q']'$).

3 Our methodology

This section presents the main components of the approach we propose to approximate the nonparametric frontiers described above. Section "Best approximation and estimation" describes the estimation of the best approximation in the proposed two-stage approach; Section "Local linear approximation" shows how to apply in the second stage the local linear approximation that allows any continuous and differentiable function to be approximated; Section "Derivatives: from factors to original units" shows how to derive the coefficients in original units when input or output factors are used; Section "Dealing with environmental factors" shows how to include environmental/external factors in this framework, and finally, Section "Bootstrap-based practical inference" describes a bootstrap-based approach to derive confidence intervals on the estimated coefficients.

3.1 Best approximation and estimation

The problem is to estimate the function $\phi(\cdot)$ in (2.13) 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 (2.13) for convenience

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

This is exactly the paradigm described in FS but in the transformed space. We might be tempted to use classical regression techniques to estimate $\phi(\cdot)$, Let $\mu_\delta(v)=\mathbb{E}(\delta(V)|V=v)$, 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

$$U_i=r_1(V_i)-\varepsilon_i, \tag{3.1}$$

where $\varepsilon_i=||d||(\delta_i-\mu_\delta(V_i))$ so that now $\mathbb{E}(\varepsilon_i|V_i)=0$. 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. For the one parameter scale family (like Exponential or Half Normal), knowledge of the variance is enough. This variance may be estimated by regressing in a second stage the squared residuals from the regression in (3.1) on v. Then we can derive $\widehat{\mu}_\delta(v)$ and shift back $\widehat{r}_1(v)$ to get the estimator $\widehat{\phi}(v)$. 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 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 preconditions the characteristic of the final estimate of ϕ. In particular a wrong choice provides unreliable 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 fully 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.

The extension of FS's approach to our framework goes along the following lines. Since the frontier function $\phi(v)$ is unknown, we consider as a starting point a class of parametric functions that can be written as $\{\phi_\theta|\theta\in\mathbb{R}^k\}$, 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 $\{\phi_\theta|\theta\in\mathbb{R}^k\}$ is defined through the pseudo-true value of θ:

$$\theta_0=\arg\min_{\theta\in\mathbb{R}^k}\int(\phi(v)-\phi_\theta(v))^2f_V(v)dv. \tag{3.2}$$

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

$$\begin{aligned}\theta_{0,n} &= \arg\min_{\theta \in \mathbb{R}^k} \textstyle\sum_{i=1}^n [\phi(V_i) - \phi_\theta(V_i)]^2, \\ &= \arg\min_{\theta \in \mathbb{R}^k} \textstyle\sum_{i=1}^n \left[U_i^\partial - \phi_\theta(V_i)\right]^2,\end{aligned} \tag{3.3}$$

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[8]

$$\begin{bmatrix} V_i \\ U_i \end{bmatrix} = R_d \begin{bmatrix} X_i \\ Y_i \end{bmatrix}, \tag{3.4}$$

where R_d is the fixed nonrandom matrix defined in (2.4).

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

$$\widehat{U}_i^\partial = U_i + ||d|| \widehat{\delta}_i, \tag{3.5}$$

which are the nonparametric estimates of the unobserved true values

$$U_i^\partial = U_i + ||d|| \delta_i. \tag{3.6}$$

[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:

$$\widehat{\theta}_n = \arg\min_{\theta \in \mathbb{R}^k} \sum_{i=1}^n \left[\widehat{U}_i^\partial - \phi_\theta(V_i)\right]^2, \tag{3.7}$$

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 also gives estimates of its derivatives $\widehat{\nabla \phi}(v) = \widehat{\partial \phi(v)} / \partial v$ and we know from Section "Some relations between the two spaces" how to recover, from these estimates, the objects of interest in the original $w = (x, y)$-space.

For interested readers, Appendix B in the SM presents the extension of the linear approximation of FS to the multivariate case and with directional distances.

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

3.2 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. This is why we propose to smooth the frontier by flexible local linear models.

In place of looking for the best linear approximation, as done in FS, 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 through order 2), we can use the first-order terms of a Taylor expansion of the function around v

$$\phi(\tilde{v}) = \phi(v) + \left(\frac{\partial \phi(\tilde{v})}{\partial v}\right)'_{v = v} (\tilde{v} - v) + o(||\tilde{v} - v||), \tag{3.8}$$

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

$$\phi(\tilde{v}) = \alpha(v) + \beta'(v)(\tilde{v} - v). \tag{3.9}$$

Here, in the spirit of (3.2), 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

$$(\alpha_0(v), \beta_0(v)) = \arg\min_{\alpha,\beta} \int (\phi(\tilde{v}) - [\alpha + \beta'(\tilde{v} - v)])^2 K_h(\tilde{v} - v) f_V(\tilde{v}) d\tilde{v}, \tag{3.10}$$

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

$$K_h(\tilde{v} - v) = \prod_{j=1}^{r-1} \frac{1}{h_j} K\left(\frac{\tilde{v}_j - v_j}{h_j}\right), \tag{3.11}$$

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 (3.3) as

$$(\alpha_{0,n}(v), \beta_{0,n}(v)) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} \left(U_i^{\partial} - [\alpha + \beta'(V_i - v)]\right)^2 K_h(V_i - v), \tag{3.12}$$

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 (3.12). It is known (see e.g. Fan and Gijbels 1996) that

$$\theta_{0,n}(v) = \begin{pmatrix} \alpha_{0,n}(v) \\ \beta_{0,n}(v) \end{pmatrix} = (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{U}^{\partial}, \tag{3.13}$$

where the $n \times (p+q)$ matrix $\mathcal{V}$ has its ith row given by $[1 \ \ (V_i - v)']$ 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 on the frontier, i.e. $\phi(V_i)$.

For the order-m frontier, we have

$$\theta_{0,n}^{(m)}(v) = \begin{pmatrix} \alpha_{0,n}^{(m)}(v) \\ \beta_{0,n}^{(m)}(v) \end{pmatrix} = (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{U}_m^{\partial}, \tag{3.14}$$

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 through order 2) we know by (3.8) that $\alpha_{0,n}(v) \to \phi(v)$ and $\beta_{0,n}(v) \longrightarrow \partial\phi(\tilde{v})/\partial v$ and that $\alpha_{0,n}^{(m)}(v) \longrightarrow \phi_m(v)$ and $\beta_{0,n}^{(m)}(v) \longrightarrow \partial\phi_m(\tilde{v})/\partial v$ 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 $\theta(v) = [\alpha(v) \ \beta'(v)]'$ will be estimated from the sample $\{V_i, \widehat{U}_i^{\partial}\}_{i=1}^{n}$ by the weighted constrained least squares problem

$$(\widehat{\alpha}_n(v), \widehat{\beta}_n(v)) = \arg\min_{\alpha,\beta} \left[\sum_{i=1}^{n} \left(\widehat{U}_i^{\partial} - (\alpha + \beta'(V_i - v))\right)^2 K_h(V_i - v)\right]. \tag{3.15}$$

From the Taylor expansion (3.8), 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

$$\widehat{\theta}_n(v) = \begin{pmatrix} \widehat{\alpha}_n(v) \\ \widehat{\beta}_n(v) \end{pmatrix} = (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\widehat{\boldsymbol{U}}^{\partial}. \tag{3.16}$$

Clearly we have for all v,

$$\begin{aligned}\widehat{\theta}_n(v) - \theta_{0,n}(v) &= (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\left(\widehat{\boldsymbol{U}}^{\partial} - \boldsymbol{U}^{\partial}\right) \\ &= \|d\|(\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\left(\widehat{\boldsymbol{\delta}}^{\partial} - \boldsymbol{\delta}^{\partial}\right),\end{aligned} \tag{3.17}$$

which is a locally weighted version of what we have for the simple linear approximations (see Appendix B in SM). In practice the bandwidths h are determined by least-squares cross validation (LSCV) and this provides bandwidths with an optimal order $h_j = O\left(n^{-1/(r+3)}\right)$ since $v \in \mathbb{R}^{r-1}$. Since for a given v, (3.17) is a simple linear transformation of the estimation errors $\widehat{\boldsymbol{\delta}}^{\partial} - \boldsymbol{\delta}^{\partial}$, we keep the same properties as described in FS, 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

$$\widehat{\theta}_n^{(m)}(v) = \begin{pmatrix} \widehat{\alpha}_n^{(m)}(v) \\ \widehat{\beta}_n^{(m)}(v) \end{pmatrix} = (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\widehat{\boldsymbol{U}}_m^{\partial}. \tag{3.18}$$

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

$$\begin{aligned}\widehat{\theta}_n^{(m)}(v) - \theta_{0,n}^{(m)}(v) &= (\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\left(\widehat{\boldsymbol{U}}_m^{\partial} - \boldsymbol{U}_m^{\partial}\right) \\ &= \|d\|(\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\boldsymbol{\mathcal{V}})^{-1}\boldsymbol{\mathcal{V}}'\boldsymbol{W}(v)\left(\widehat{\boldsymbol{\delta}}_m^{\partial} - \boldsymbol{\delta}_m^{\partial}\right),\end{aligned} \tag{3.19}$$

where here again the bandwidths can be selected by LSCV. So for a given v we keep the properties established in Theorem 3.1 of FS, specifically, for any v,

$$\sqrt{n}\left(\widehat{\theta}_n^{(m)}(v) - \theta_{0,n}^{(m)(v)}\right) \xrightarrow{\mathcal{L}} N(0, \Sigma(v)), \tag{3.20}$$

where $\Sigma(v)$ is a matrix depending on several characteristics of the DGP.

3.3 Derivatives: from factors to original units

The “curse of dimensionality” problem is well known in the field of nonparametric efficiency analysis and consists of the need to use large datasets in order to have estimates with an acceptable level of precision (see e.g., Daraio and Simar 2007). Therefore, it is important in this area to reduce the dimensionality of the analysis by aggregating inputs or outputs into factors if they are highly correlated (Wilson 2018). For this reason, we remind below how to build input or output factors and we describe how to recover the derivatives in original units from those obtained from the factors.

Consider an output factor F_y defined as $F_y = a'\check{y}$ for some $\check{y} \in \mathbb{R}_+^L$ 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 $y = (\check{y}' \ \mathring{y}')'$. We know that $a \in \mathbb{R}^L$, with $a'a = 1$ is the eigenvector of the 2nd moments matrix of the L outputs $\check{y}$, corresponding to its largest eigenvalue.

Similarly we may have an input factor F_x defined as $F_x = b'\check{x}$ for some $\check{x} \in \mathbb{R}^J$ 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 $\check{x}$, 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 $x = (\check{x}' \ \mathring{x}')'_1$.

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 $\check{y}$-space given by

$$\check{y} = F_y a \in \mathbb{R}^L_+. \tag{3.21}$$

Consider now a differential ∂F_y relative to some other variables. This differential shifts the frontier point along the direction a at the point $F_y + \partial F_y$, which has coordinates in the $\check{y}$-space given by $(F_y + \partial F_y)a = \check{y} + \partial F_y a$, so that we may define the corresponding differential in the $\check{y}$-space as

$$\partial\check{y} = \partial F_y a \in \mathbb{R}^L. \tag{3.22}$$

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

$$\partial\check{x} = \partial F_x b \in \mathbb{R}^J. \tag{3.23}$$

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 $\mathring{x}$

$$\frac{\partial\check{y}}{\partial\mathring{x}'} = \frac{\partial F_y a}{\partial\mathring{x}'} = a\frac{\partial F_y}{\partial\mathring{x}'}, \tag{3.24}$$

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

$$\frac{\partial\mathring{y}}{\partial\check{x}'} = \frac{\partial\mathring{y}}{\partial F_x b'} = \frac{\partial\mathring{y}}{\partial F_x}(i_J \oslash b)', \tag{3.25}$$

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 $(\partial\mathring{y}_k/\partial F_x)(1/b_j)$.

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

$$\frac{\partial\check{y}}{\partial\check{x}'} = \frac{\partial F_y a}{\partial F_x b'} = \frac{\partial F_y}{\partial F_x}a(i_J \oslash b)'. \tag{3.26}$$

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

$$\mathcal{E}(\check{y}_\ell, \check{x}_j) = \mathcal{E}(F_y, F_x), \tag{3.27}$$

or, for another instance

$$\mathcal{E}(\check{y}_\ell, \mathring{x}_i) = \mathcal{E}(F_y, \mathring{x}_i), \tag{3.28}$$

and many other possibilities.

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

$$\frac{\partial\check{y}}{\partial\mathring{y}'} = \frac{\partial F_y a}{\partial\mathring{y}'} = a\frac{\partial F_y}{\partial\mathring{y}'}, \tag{3.29}$$

i.e. a $(L \times K)$ matrix with element (ℓ, k) given by $a_\ell(\partial F_y/\partial\mathring{y}_k)$.[9]

3.4 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, 2007) have introduced the concepts of conditional frontiers and conditional efficiency measures. This leads to define a conditional attainable set

$$\Psi^z = \{(x, y) | x \text{ can produce } y \text{ when } Z = z\}. \tag{3.30}$$

Clearly $\Psi^z \subset \Psi$ which includes as a 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

[9] Note that this approach does not allow to recover from the factors the marginal rates of substitution between the elements $\check{x}$ composing the factor F_x, or the marginal rates of transformation between the outputs $\check{y}$ composing the factor F_y. For instance the information about $\partial\check{y}_{\ell_1}/\partial\check{y}_{\ell_2}$ 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.

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

$$\delta(x,y|z) = \sup\{\delta |\, (x - \delta d_x, y + \delta d_y) \in \Psi^z\}. \tag{3.31}$$

Nonparametric estimators have been proposed by Cazals et al. (2002) and Daraio and Simar (2005) and most of their statistical properties are derived in Jeong et al. (2010). CLTs have been obtained for averages of these measures in Daraio et al. (2018) and practical Matlab programs for computation are reported in Daraio et al. (2020).

It is useful to point out that none of these previous works have proposed approximations of the nonparametric efficient frontiers and related coefficients.

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 in 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 $\delta(x, y|z)$ and their estimators. The transformation of w to t in (2.5) remains the same providing, by (2.9), 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 (3.5)

$$\widehat{U}^{\partial}(Z_i) = U_i + ||d||\widehat{\delta}(W_i|Z_i), \tag{3.32}$$

and the approximating Eq. (3.7) becomes

$$\widehat{\phi}(\cdot,\cdot) = \arg\min_{\phi(\cdot,\cdot)} \sum_{i=1}^{n} \left[\widehat{U}^{\partial}(Z_i) - \phi(V_i, Z_i)\right]^2, \tag{3.33}$$

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, $\sqrt{n}$ becomes $n^{2/(r_z+4)}$, where r_z is the dimension of Z.

3.4.1 Recovering derivatives in full space including z variables

Although this is rather obvious, it is better to clarify how the relations between the two spaces described above from (2.14) to (2.18), have to be adapted to the full space, including the variables z, where the values $\tilde{w} = (w, z)$ are transformed in $\tilde{t} = (t, z)$. We have now

$$\tilde{t} = \widetilde{R}_d \tilde{w} = \begin{bmatrix} R_d & 0 \\ 0 & I_{r_z} \end{bmatrix} \begin{pmatrix} w \\ z \end{pmatrix}, \tag{3.34}$$

since the rotation is only done on the w variables. The inverse transformation can be written as

$$\tilde{w} = \widetilde{R}_d' \tilde{t} = \begin{bmatrix} R_d' & 0 \\ 0 & I_{r_z} \end{bmatrix} \begin{pmatrix} t \\ z \end{pmatrix}. \tag{3.35}$$

By following the same argument as in Section "Some relations between the two spaces", in the $\tilde{t}$ space, the hyperplane tangent at the frontier $\phi(v, z)$ at the frontier point $\tilde{t}_0^{\partial} = (v_0' \; \phi(v_0, z_0) \; z_0')'$ is given by the equation $c_{\tilde{t}}'(\tilde{t} - \tilde{t}_0^{\partial})$ where now

$$c_{\tilde{t}} = \begin{bmatrix} \nabla_v \phi(v_0, z_0) \\ -1 \\ \nabla_z \phi(v_0, z_0) \end{bmatrix}, \tag{3.36}$$

where $\nabla_v \phi(v_0, z_0) = \left[\frac{\partial \phi(v,z)}{\partial v}\right]_{v=v_0, z=z_0}$ is the $(r-1)$-vector of the partial derivatives of $\phi(v, z)$ wrt v, evaluated at (v_0, z_0) and similarly, $\nabla_z \phi(v_0, z_0) = \left[\frac{\partial \phi(v,z)}{\partial z}\right]_{v=v_0, z=z_0}$ represents the r_z-vector of the partial derivatives of $\phi(v, z)$ wrt z, evaluated at the same point.

In the $\tilde{w}$ space, this hyperplane has the equation $c_{\tilde{t}}' \widetilde{R}_d (\tilde{w} - \tilde{w}_0^{\partial}) = 0$ or

$$c_{\tilde{w}}'(\tilde{w} - \tilde{w}_0^{\partial}) = 0, \text{ where } c_{\tilde{w}} = \widetilde{R}_d' c_{\tilde{t}} = \begin{bmatrix} R_d' \begin{bmatrix} \nabla_v \phi(v_0, z_0) \\ -1 \end{bmatrix} \\ \nabla_z \phi(v_0, z_0) \end{bmatrix}. \tag{3.37}$$

The last equation gives an explicit expression for any partial derivatives at frontier points $\tilde{w}_0^{\partial}$ in the original units. For instance we have, as Eq. (2.17) above in Section " Some relations between the two spaces"

$$\left.\frac{\partial \tilde{w}_{\ell}}{\partial \tilde{w}_k}\right|_{\tilde{w} = \tilde{w}_0^{\partial}} = -\frac{c_{\tilde{w},k}}{c_{\tilde{w},\ell}}, \tag{3.38}$$

provided $c_{\tilde{w},\ell} \neq 0$. This allows us to recover all the characteristics of the frontier at any frontier points by selecting the appropriate elements of $\tilde{w}$. This includes the (conditional to z) marginal rates of substitution and the (conditional to z) marginal rates of transformation. This is a useful tool for the practitioner to investigate the effect of z

on these rates and on the shape of the frontier. This will be illustrated in the real data example below.

3.5 Bootstrap-based practical inference

Practical inference in this setting will be obtained with the bootstrap.

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 (2.1) 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 randomly 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 of Simar and Wilson 2008 or in Section 3.1.2 of 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 $i = 1, \ldots, n$.

For the order-m case, 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 (3.18) in the bootstrap world, keeping the same matrices $\boldsymbol{\mathcal{V}}$ and $\boldsymbol{\mathcal{W}}(v)$, 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 (3.20).

The bootstrap also provides the bootstrap values of the quantities of interest described in Section "Some relations between the two spaces", 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.

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 $\widehat{\delta}_m^*(W_i|Z_i)$ of $\widehat{\delta}_m(W_i|Z_i)$, 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

$$\widehat{U}_m^{\partial,*}(Z_i) = U_i + ||d||\widehat{\delta}_m^*(W_i|Z_i). \tag{3.39}$$

We apply this bootstrap-based approach to simulated data in Appendices B.4 and C of Supplementary Materials. The next section reports an illustration of our approach with a real dataset.

4 Application on European universities

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, and the choice of the functional form. Subsequently, Figlio (1999), Dewey et al. (2000) and Baker (2001) expanded the investigation of the results obtained from education production functions (based on parametric approaches), showing that the imposition of restrictive assumptions leads to different results.

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 pioneering project AQUAMETH described in Bonaccorsi and Daraio (2007), in which the first empirical evidence at the European comparative level is reported. 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 et al. 2017).

It is important to note that before the introduction of our method, none of the tables of results presented in this section and none of the figures would have been possible. Prior to our method, in the nonparametric approach (DEA, FDH, order-m) it was only possible to estimate efficiency scores with their respective confidence intervals. With our method, it is now possible to complement the results provided by the nonparametric approach with the estimation of *local* coefficients of economic interest, including partial derivatives of outputs concerning specific inputs or external variables or marginal products, and by providing confidence intervals on these coefficients that were not available before.

We illustrate our methodology by analyzing a sample of 337 observations from European universities that was recently analysed in Daraio et al. (2021) to which readers are referred for more details and information. In particular, we will show that with our approach we are able to obtain results that were not available before which are useful

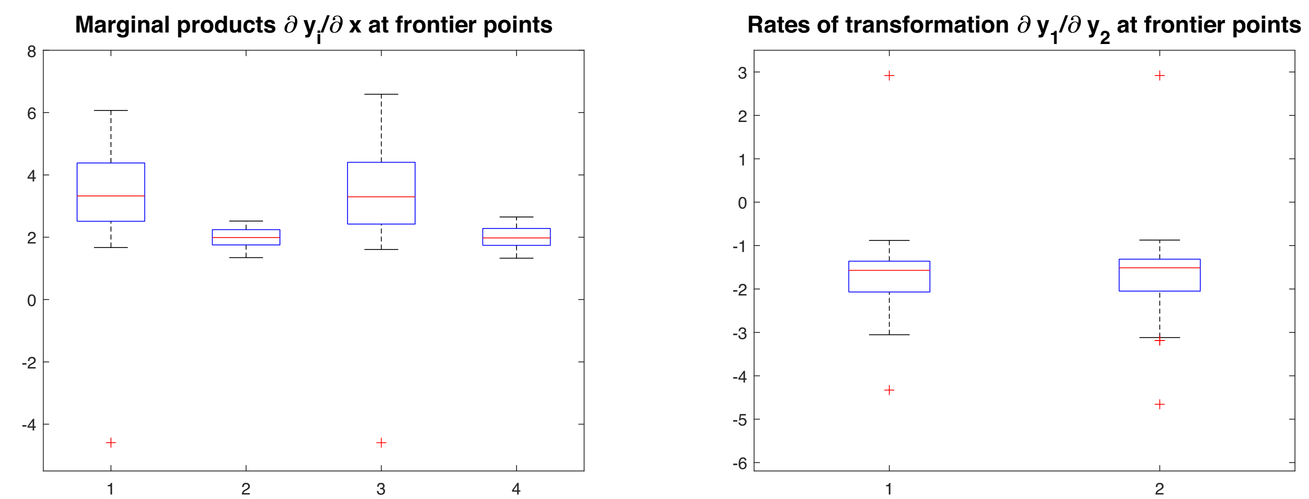

Fig. 2 HEI data: distribution of the estimated marginal and transformation rates at frontier points. Left panel, marginal rates: FDH case $\partial y_1/\partial x$ (boxplot 1) and $\partial y_2/\partial x$ (boxplot 2), order-m case with $m = 550$: $\partial y_1/\partial x$ (boxplot 3) and $\partial y_2/\partial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

information to complement those usually obtained within the nonparametric frontier framework.

We use as input an input factor (FX) that aggregates three inputs (total number of academic staff, total number of non-academic staff and and total expenditures), two outputs $Y = (TDEG, FY)$ the first one being the teaching activity (total number of degrees) and the second an output factor summarizing the research activity, being the aggregation of number of PhD students and total publications. As described in detail in Daraio et al. (2021), due to the high correlations among the 3 inputs, higher than 90%, we do not lose much information by aggregating them into one input factor.[10]Wilson (2018) has shown the advantage of such dimension reduction methods for nonparametric models of production. Moreover, in case partial derivatives involving each original input are wanted, in Section "Derivatives: from factors to original units" we described the way to recover these from the derivatives on the factor. As environmental variables we consider the latent factor quality (QUAL) identified in Daraio et al. (2021) and the specialization index (SPEC) varying between 0 and 1 and indicating, respectively, generalist versus specialist universities.

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 $scZ = (0.2903, 0.1249)$.

As often chosen in this literature (see e.g. Bonaccorsi and Daraio 2007), we select as directional vector the vector determined by the median of the input factor (with negative sign) and the median of the outputs. This gives $d = (-0.87691.19850.6679)'$, hence all three variables are active in the estimation of the optimal frontier.

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 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 $\widehat{\delta}_m(X_i, Y_i|Z_i)$. Additional details on how to select the value of m are available in Daraio and Simar (2007).

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

To illustrate the results obtained with our methodology, 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 2 in the left panel, shows the distribution of the estimated marginal products reporting the FDH case in boxplots 1 and 2 and the order-m case with $m = 550$ in

[10] Daraio et al. (2021) provide the first eigenvector of the moment matrix of the 3 original inputs: $b = (0.5723, 0.6218, 0.5346)'$, indicating that the input factor is roughly the average of the scaled inputs. This first factor explains 96% of the total inertia and it as a correlation with the 3 original inputs respectively equal to 0.9777, 0.9474 and 0.9325.

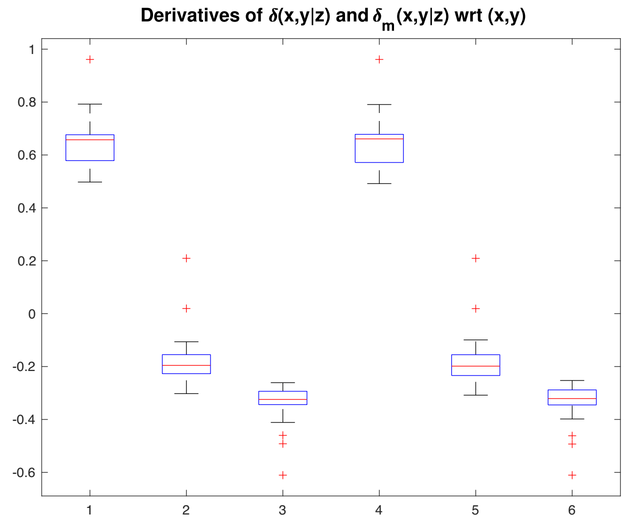

Fig. 3 HEI data: distribution of the estimated derivatives of the distance functions with respect to x, y_1 and y_2. On the left, FDH case: estimated derivatives of $\delta(x, y|z)$ with respect to x (boxplot 1), w.r.t. y_1 (boxplot 2) and w.r.t. y_2 (boxplot 3). On the right, order-m case: estimated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

boxplots 3 and 4. We observe that the distributions of the marginal products $\partial y_1/\partial x$ and $\partial y_2/\partial x$ estimated with FDH and order-m are globally the same. The marginal product $\partial y_1/\partial x$ measures the rate of change in the teaching activity of the analyzed universities (y_1 is *TDEG*) as the input changes (x is the input factor *FX*). $\partial y_2/\partial x$ measures the change in research activity (y_2 is *FY*) as input changes. Figure 2 shows that teaching activity has greater sensitivity to varying input than research activity (the distribution of $\partial y_1/\partial x$ is greater than that of $\partial y_2/\partial x$).

Figure 2, right panel, illustrates the boxplots of the transformation rates ($\partial y_1/\partial y_2$) estimated with FDH (boxplot 1) and order-m with $m = 550$ (boxplot 2). Again, the two distributions look the same. We note that the transformation rates between the two outputs (y_1 the teaching output and y_2 the research output) globally are negative and thus have the expected sign. This means that universities must strategically address a trade-off between teaching and research in their allocation of resources.

Using our directional distance frontier approximation approach, we can calculate the sensitivity of directional conditional distances to changes in the input (x) and outputs of teaching (y_1) and research (y_2). Figure 3 shows the boxplots of the estimated derivatives of the conditional 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). We observe again that the derivatives have the expected signs in that inefficiency increases as inputs increase (the derivative with respect to x_1 is positive) while inefficiency decreases when y_1 and y_2 increase. We also note that increasing research output (y_2) has a greater impact on reducing inefficiency than teaching output (y_1) because the distribution of derivatives with respect to y_2 (boxplots 3 and 6 of Fig. 3) has a median around -0.35 while the distribution of derivatives with respect to y_1 (boxplots 2 and 5 of Fig. 3) has a median around -0.2.

Figure 4 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). The plots shown in Fig. 4 show that there is great heterogeneity among the units analyzed. While the plot on the left side of the figure shows some decreasing trend in the variation of inefficiency with respect to x (it would seem that universities with higher inputs increase inefficiency less than those with lower inputs); the middle and right plots show more heterogeneity in the trend of inefficiency with respect to y_1 and y_2 outputs.

All the figures displayed so far show a great variability. This is not a surprise since we have estimates at each data points and our flexible procedure allows for heteroskedasticity. Hence, the figures and various boxplots show the full distribution of the estimates over the whole data set where the environmental factors changes from one point to another. In a final set of figures we will investigate the effect of the z-variables on the input and the outputs and on the shape of the frontier. To save space we only display the figures for the full frontier estimates (we have similar pictures for the order-m frontier case).

First we can compute, by Eq. (3.38) the partial derivatives of the input and the outputs with respect to the two variables z (latent quality *QUAL* and specialization *SPEC*). Figure 5 displays the full distribution of these 337 partial derivatives. From Fig. 5 we see that the two variables z act as an “output”: they show mainly positive derivatives with respect to x and negative derivatives with respect to both outputs. In addition, we observe that the impact of specialization (*SPEC*) on x is smaller in magnitude (showing a median close to zero) than the impact of quality (*QUAL*) (with a median of about 0.1), see the left plot in Fig. 5. The impact of *SPEC* is also smaller than the impact of *QUAL* on teaching (y_1) and research (y_2) outputs as shown in the middle and right plots of Fig. 5. To better investigate the dependence of the shape of the frontier, described by the partial derivatives of the outputs with respect to the input, as a function of z we show Fig. 6. Again these partial derivatives are computed by (3.38) at each data point having its own value of z. Figure 6 displays a local linear fit of the resulting cloud of points. We see as expected that the variable *SPEC* has less effect than the variable *QUAL*. The latter decreases the curvature in the direction of y_1 (teaching activity) when it increases showing that universities do not

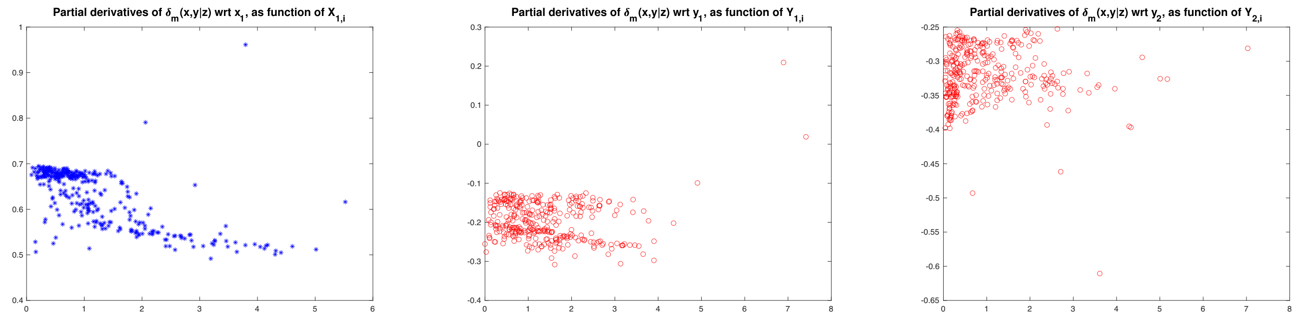

Fig. 4 HEI data: plots of the estimated derivatives of $\delta_m(x, y|z)$, from left to right with respect to x, y_1 and y_2 versus the observed values of x, y_1 and y_2

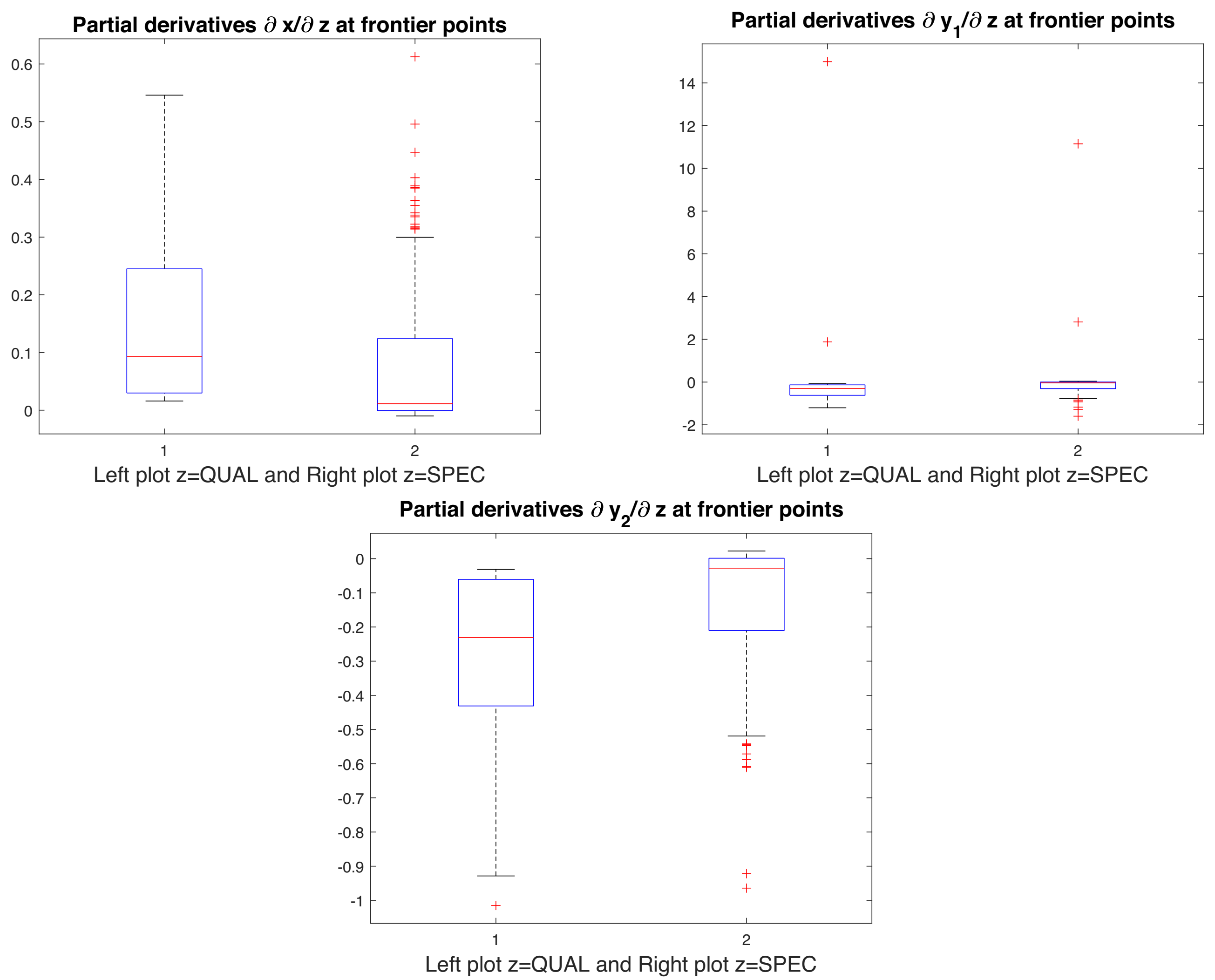

Fig. 5 HEI data: From left to right, plots of the estimated derivatives of $\partial x/\partial z$, $\partial y_1/\partial z$ and $\partial y_2/\partial z$ (in each plot, left case $z_1 = QUAL$ and right case $z_2 = SPEC$)

have to produce a lot of teaching if they have a high level of quality in that QUAL acts as a kind of compensatory “output". Interestingly, the effect of z on the curvature of the frontier in the direction of the research output y_2 (Fig. 6 right panel) looks like an inverted U-shaped. The impact of quality on research output (y_2) is more complex and not as monotone decreasing as that on teaching (y_1). As Fig. 6 in the right panel shows, for research activity, the

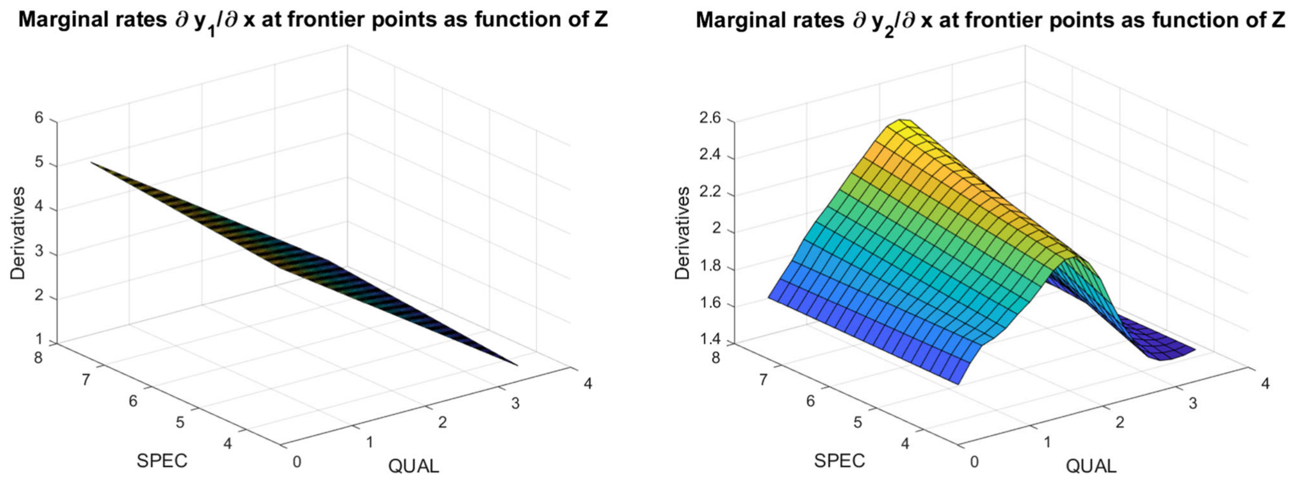

Fig. 6 HEI data: Local linear fit of the estimated derivatives of $\partial y/\partial x$, as a function of z, left panel y_1 and right panel y_2

Table 1 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 $m = 550$

	$\partial y_1/\partial x_1$		$\partial y_2/\partial x_1$		$\partial y_1/\partial y_2$	
	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
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

“substitution" or compensatory effect of quality (allowing for a reduction in output y_2) begins only after a certain level of quality is reached, the one where the decreasing part of the inverted U-shaped curve begins. This shows us that the effect of the $QUAL$ (z_2) variable is neither uniform nor constant, but changes depending on the observations. Finally, note also that the left panel of Fig. 6 does not indicate interaction effects between the two variables z. On the contrary, in the right panel of Fig. 6 a small interaction between the z appears: here specialization ($SPEC$) seems to enhance the impact of x on research (y_2) when quality ($QUAL$) increases.

Another important contribution of our approach is the possibility of estimating the statistical significance of coefficients approximating nonparametric frontiers. As we described in Section “Bootstrap-based practical inference”, by applying bootstrapping in our framework we can obtain confidence intervals on each estimated parameter. Table 1 shows the 95% bootstrap confidence intervals for the rates $\partial y_1/\partial x_1$ (second and third columns), $\partial y_2/\partial x_1$ (fourth and fifth

Table 2 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 $m = 550$

	$\partial\delta_m(x, y\|z)/\partial x$		$\partial\delta_m(x, y\|z)/\partial y_1$		$\partial\delta_m(x, y\|z)/\partial y_2$	
	lower bound	upper bound	lower bound	upper bound	lower bound	upper bound
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

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% level except for units #9, 15, 16, 18 and 19 which include zero in their confidence intervals.

Table 2 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% level. Overall, we observe that the estimated quantities showed in Tables 1 and 2 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 Section "Derivatives: from factors to original units", we described how these derivatives in "factor units" can provide the derivatives in the original units.[11]

[11] Just to illustrate this, we could recover, e.g., the partial derivatives of any output, say y_1 at the frontier points wrt to the original inputs by multiplying the derivative wrt the the input factor by the eigenvector b given in Footnote 10. This is obtained by applying equation (3.25) in Section "Derivatives: from factors to original units".

In practice more detailed analysis could be done, however, the illustrative example reported in this section clearly shows the usefulness and the flexibility of our approach in complementing the analysis available within a complete nonparametric framework.

5 Conclusions

Nonparametric methods that provide envelopment estimators, such as FDH or DEA, are very attractive as they do not rely on restrictive parametric assumptions on the DGP, specifically on the shape of the boundary and on the distribution of inefficiency. However, these nonparametric techniques do not allow us to make sensitivity analyses of the estimated frontiers, for example, to estimate derivatives of the optimal production outputs concerning specific inputs or infer marginal products and other coefficients of the frontier of economic interest.

In this paper, we propose an approach that complements and completes existing nonparametric efficiency methods by providing approximations of the economic coefficients of interest by "smoothing" the nonparametric estimators of the frontiers. It is an extension and generalization of the ideas initiated by Florens and Simar (2005), who propose linear models to approximate univariate FDH or order-m frontier functions.

This work represents an important step forward in the field of nonparametric efficiency analysis. To the best of our knowledge, it provides a unique approach that has not been proposed so far in the literature. It offers the possibility of nonparametrically approximating all desired coefficients and partial derivatives, with their bootstrap-estimated confidence intervals, in a fully multivariate directional distance model that includes environmental factors. In detail, the novelty of our approach is manifold. It allows us to handle fully *multivariate* cases in a flexible *directional distance* model. It provides flexible approximations based on *local linear* tools offering local estimates of all the desired coefficients and partial derivatives without assuming any parametric structure. It proposes simple *bootstrap algorithms* to estimate confidence intervals on all the coefficients of interest. It extends the method for including environmental factors and estimating their impact in this framework. Illustrations with some simulated data sets and with real data show the usefulness and flexibility of the proposed approach. In particular, the application on European universities shows the wealth of economic coefficients and derivatives estimated nonparametrically, made available by this approach.

This approach has a high potential for applicability in many different contexts. One of these is the field of regulated sectors in which policymakers need *coefficients* for their economic interpretation, for setting their price-cap, and to monitor the efficiency of regulated industries. Thanks to our approach, they are not forced to rely on very strict and unrealistic production function specifications for estimating the efficient frontier from which they derive the economic coefficients. In this sense, our approach avoids the empirical choice between a parametric and nonparametric approach because it offers to those that use the nonparametric approach the availability of coefficients estimated nonparametrically that do not rely on restrictive assumptions and are available for each firm, institution, or observation in the sample.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s11123-024-00726-2.

Acknowledgements The financial support of Sapienza project awards n. RM12117A8A5DBD18 is gratefully acknowledged.

Author contributions All authors contributed equally to the paper.

Funding Open access funding provided by Università degli Studi di Roma La Sapienza within the CRUI-CARE Agreement.

Conflict of interest The authors declare no competing interests.

References

Baker BD (2001) Can flexible non-linear modeling tell us anything new about educational productivity? Econ Educ Rev 20:81–92

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

Cazals C, Florens JP, Simar L (2002) Nonparametric frontier estimation: a robust approach. J Econ 106:1–25

Chambers RG, Chung Y, Färe R (1998) Profit, directional distance functions, and nerlovian efficiency. J Optim Theory Appl 98:351–364

Daouia A, Florens JP, Simar L (2012) Regularization of nonparametric frontiers estimators. J Econ 168:285–299

Daraio C, Simar L (2005) Introducing environmental variables in nonparametric frontier models: a probabilistic approach. J Product Anal 24:93–121

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

Daraio C, Simar L, Wilson PW (2018) Central limit theorems for conditional efficiency measures and tests of the “separability" condition in nonparametric, two-stage models of production. Econ J 21:170–191

Daraio C, Simar L, Wilson PW (2020) Fast and efficient computations of directional distances estimators. Ann Oper Res 288:805–835

Daraio C, Simar L, Wilson PW (2021) Quality as a latent heterogeneity factor in the efficiency of universities. Econ Model 99:105485. *Econometrica*, 19:3,273-292

De Witte K, Lopez-Torres L (2017) Efficiency in education: a review of literature and a way forward. J Oper Res Soc 68:339–363

Dewey J, Husted TA, Kenny LW (2000) The ineffectiveness of school inputs: a product of misspecification? Econ Educ Rev 19:27–45

Färe R, Grosskopf S, Lovell CAK (1985), *The Measurement of Efficiency of Production*. Boston, Kluwer-Nijhoff Publishing.

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

Figlio DN (1999) Functional form and the estimated effects of school resources. Econ Educ Rev 18:241–252

Florens JP, Simar L (2005) Parametric approximations of nonparametric frontier. J Econ 124:91–116

Grosskopf S, Hayes K, Taylor LL (2014) Applied efficiency analysis in education. Econ Bus Lett 3:19–26

Hafner CM, Manner H, Simar L (2018) The “wrong skewness” problem in stochastic frontier models: a new approach. Econom Rev 37:380–400

Hanushek EA (1979) Conceptual and empirical issues in the estimation of educational production functions. J Human Resour 14:351–388

Jeong SO, Park BU, Simar L (2010) Nonparametric conditional efficiency measures: asymptotic properties. Ann Oper Res 173:105–122

Jeong SO, Simar L (2006) Linearly interpolated FDH efficiency score for nonconvex frontiers. J Multivar Anal 97:2141–2161

Johnes J (2006) Data envelopment analysis and its application to the measurement of efficiency in higher education. Econ Educ Rev 25:273–288

Kneip A, Simar L, Wilson PW (2008) Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models. Econom Theory 24:1663–1697

Kneip A, Simar L, Wilson PW (2015) When bias kills the variance: Central Limit Theorems for DEA and FDH efficiency scores. Econom Theory 31:394–422

Kuosmanen T, Johnson A (2017) Modeling joint production of multiple outputs in StoNED: Directional distance function approach. Eur J Oper Res 262:792–801

Park B, Simar L, Weiner CH (2000) The FDH estimator for productivity efficiency scores: asymptotic properties. Econom Theory 16:855–877

Podinovski V (2019) Direct estimation of marginal characteristics of nonparametric production frontiers in the presence of undesirable outputs. Eur J Oper Res 279:258–276

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

Shephard RW (1970). *Theory of cost and production function*. Princeton University Press, Princeton, New-Jersey.

Simar L, Vanhems A (2012) Probabilistic characterization of directional distances and their robust versions. J Econ 166:342–354

Simar L, Vanhems A, Wilson PW (2012) Statistical inference with DEA estimators of directional distances. Eur J Oper Res 220:853–864

Simar L, Van Keilegom I, Zelenyuk V (2017) Nonparametric least squares methods for stochastic frontier models. J Product Anal 47:189–204

Simar L, Wilson P (2007) Estimation and inference in two-stage, semi-parametric models of production processes. J Econ 136:31–64

Simar L, Wilson PW (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.

Simar L, Wilson PW (2013) Estimation and inference in nonparametric frontier models: recent developments and perspectives. Found Trends® Econ 5:183–337

Simar L, Wilson PW (2015) Statistical approaches for nonparametric frontier models: a guided tour. Int Stat Rev 83:77–110

Simar L, Wilson PW (2020) Hypothesis testing in nonparametric models of production using multiple sample splits. J Product Anal 53:287–303

Simar L, Wilson PW (2022), Nonparametric, stochastic frontier models with multiple inputs and outputs, J Bus Econ Stat. https://doi.org/10.1080/07350015.2022.2110882.

Stock JH (2010) The other transformation in econometric practice: Robust tools for inference. J Econ Perspect 24:83–94

Wilson PW (2018) Dimension reduction in nonparametric models of production. Eur J Oper Res 267:349–367