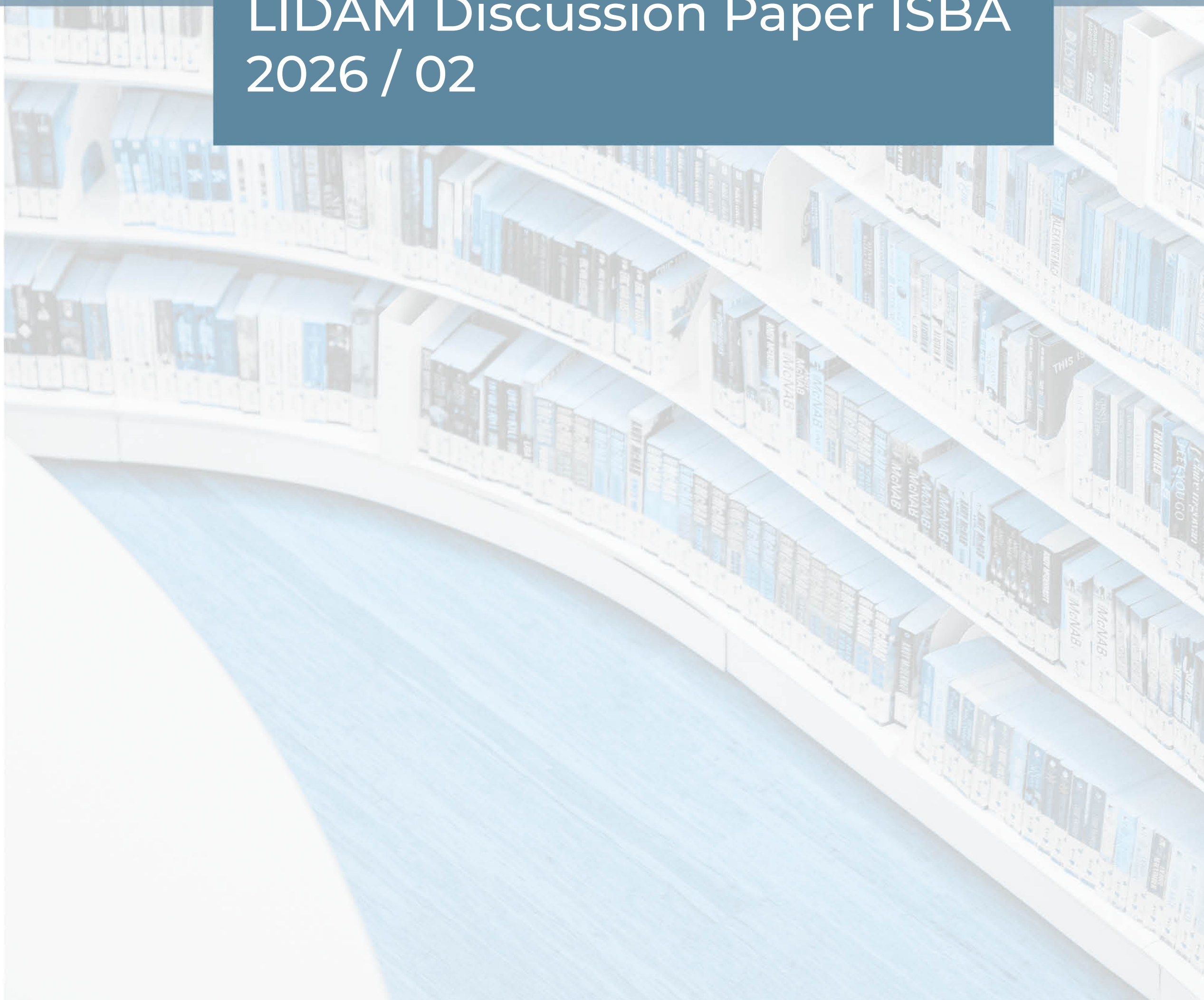
NONPARAMETRIC MODELS OF PRODUCTION: EFFICIENCY ESTIMATION AND STATISTICAL INFERENCE

Léopold Simar, Paul Wilson

LIDAM Discussion Paper ISBA
2026 / 02



UCLouvain

LIDAM
ISBA | Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA
Voie du Roman Pays 20 - L1.04.01
B-1348 Louvain-la-Neuve
Email : lidam-library@uclouvain.be
https://uclouvain.be/en/research-institutes/lidam/isba/publication

Nonparametric Models of Production: Efficiency Estimation and Statistical Inference

Léopold Simar* Paul W. Wilson§

February 2026

Abstract

Production theory is based on an economic model where we define the production set, i.e. the set of the combinations of inputs and outputs that are technically feasible. The efficiency of a particular unit is measured by its distance to the efficient frontier of the production set, based on a selected direction. Nonparametric models are particularly appealing because they do not rely on restrictive assumptions about the shape of the efficient frontier nor on the processes that may give rise to inefficiencies. Since these quantities are typically unknown, they must be estimated from a sample of observed units. The most widely used non-parametric approaches are based on envelopment estimators such as Data Envelopment Analysis (DEA) or Free Disposal Hull (FDH), making the derived measures of efficiency for a given unit dependent on these envelopment estimators. In recent decades, substantial results have been derived regarding the statistical properties of these non-parametric estimators. These advancements facilitate statistical inference regarding the efficiency scores of individual units across different contexts or efficiency comparison between groups of units, as well as testing procedures concerning the shape of the attainable set (whether convex or non-convex), or assumptions about returns to scale. It is shown how crucial the assumptions made on the DGP are, incorrect assumptions may lead to inconsistent estimators and wrong inference. These results have now been extended to dynamic settings, including inference on Malmquist Productivity Indices (and other well-known productivity indices) and their components. In this paper, we provide a comprehensive up-to-date survey of various approaches.

Key Words: Production Theory, Nonparametric estimation, Data envelopment analysis, Conditional frontiers.
JEL Classification: C1, C14, C13, D24, O47

*Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA), LIDAM, UCLouvain, Voie du Roman Pays 20, B-1438 Louvain-la-Neuve, Belgium; email: leopold.simar@uclouvain.be.

§Department of Economics, Clemson University, Clemson, South Carolina, USA; email: pww@clemson.edu.

Contents

1 Introduction

Benchmarking is an important topic in management and applied economics. The idea is to compare firms, plants, banks, hospitals, universities, etc., to other units having similar activities, but being more or less efficient in the way they combine their inputs (e.g., labor, capital, energy, etc.) to produce their outputs (e.g., goods produced, services offered, etc.). This benchmarking is useful for managers. It helps managers of less efficient units improve their unit's performance by learning from similar units that are more efficient. Benchmarking may aid in detecting and understanding why some units appear to have very low performance, or it may reveal that the researcher's model is incomplete, perhaps neglecting some important features such as heterogeneity factors.

Production theory based on the pioneering work of Koopmans (1951) and Debreu (1951) provides the elements of an economic model allowing an analysis of production. The central component of the model is the production set, i.e., the set of feasible combinations of input and output quantities. Measures of the technical efficiency of a given producer correspond to measures of distance from the point in input-output space where the producer operates to the efficient boundary of the production set. Various assumptions on the production set are commonly made, and various directions in which distance to the production set boundary is measured have been proposed. See Shephard (1970), Färe et al. (1985) or Färe (1988) for an overview.

The production set as well as other theoretical economic concepts such as the stochastic process that generates inefficiency among producers is in general unobserved. With few exceptions, the empirical researcher only observes quantities of inputs and outputs for a sample of n producers. These observations must be used to estimate technical efficiency and other features of interest. Farrell (1957) provides the first formal empirical study of technical efficiency, but without any statistical theory and thus without statistical inference. Inference is needed to quantify the uncertainty surrounding estimates, and inference requires a model of the data-generating process (DGP). The most widely-used non-parametric estimators of the production set are based on envelopment estimators, i.e., sets that envelop the observed input-output quantities in input-output space. Then estimates of efficiency are obtained as measures of distance from a particular observation (of a producer's input and output quantities) to the edge of the estimate of the production set.

Nonparametric estimators have the advantage (over parametric estimators) of being based on a minimal set of assumptions (e.g., regularity of the boundary or qualitative assumptions on the shape of the production set) and do not require, in general, specific assumptions on the process that generates inefficiency which causes producers to operate in the interior of the production set, some distance from its frontier. The price to pay for this flexibility is (i) the curse of dimensionality—as is typical in non-parametric estimation, more data are needed as dimensionality (i.e., the number of inputs and outputs) increases to avoid increasing estimation error, and (ii) the use of bootstrap techniques for making inference.

Substantial results on statistical properties of non-parametric estimators of production sets and efficiency measures have appeared in recent decades. These advances permit statistical inference on the efficiency scores of individual units as well as comparisons of efficiencies between groups of units. In addition, easily-implemented procedures for testing assumptions on the shape of the production set have been developed. As shown below, the assumptions made on the DGP, including the shape of the production set, are crucial for obtaining valid, meaningful results in empirical research. Imposition of unfounded, incorrect assumptions leads to statistically inconsistent estimates, invalid inference, and consequently, to misleading conclusions. These statistical results results have recently been extended to dynamic settings to make inferences on Malmquist Productivity Indices (MPIs) and their components. This chapter provides a comprehensive, up-to-date survey of the tools that are available to empirical researchers today.

A separate literature on parametric methods for efficiency estimation has also been developed in recent decades (e.g., see Kumbhakar and Lovell, 2000). These methods are based on particular, restrictive assumptions on the shape of the frontier and typically involve specifying a parametric function that is assumed known up to a small number of parameters to describe the frontier. In addition, almost all such models require a parametric assumption on the distribution of efficiencies. The most common specifications are the translog functional form for the frontier, and a half-normal distribution for efficiency combined with a normal distribution for noise. The main problem with these approaches is that when the specific, particular parametric assumptions are incorrect, all of the resulting inference is invalid.[1]

This chapter focuses on non-parametric methods for analysis of production performance and is organized as follows. Section 2 introduces the economic model and specifies the DGP. Section 3 presents the most popular non-parametric estimators and discusses their statistical

[1]The translog functional form is seldom tested in empirical studies of efficiency, but is often rejected when it is tested. See, for example, McAllister and McManus (1993), Wheelock and Wilson (2001, 2012) 2018). Wheelock and Wilson (2018) note that rejection of the translog functional form is hardly surprising since the translog function is merely a quadratic in log-space, which limits the variety of shapes that can be taken by the function. Wheelock and Wilson (2018) also note that the translog function is derived from a Taylor expansion of output, cost, etc. around the means of the data, and one should not expect it to fit well data that are highly variable and highly skewed, which are common features of data from many industries. See Wheelock and Wilson (2018, Section 4) for additional discussion.

properties. Section 4 presents estimation approaches that are less sensitive to outliers and extreme values than the full envelopment estimators. In Section 5 a natural way to introduce environmental variables is presented. Testing issues are addressed in Section 6. Dynamic models and MPIs are defined in Section 7, providing the tools for doing inference on gains or losses of productivity of firm across time periods. Section 8 concludes, giving a short summary of themes and extensions omitted from this chapter.

2 Nonparametric Models of Production

2.1 The Economic Model and Measures of Efficiency

The economic model for a production process where inputs $X \in \mathbb{R}_+^p$ are used to produce outputs $Y \in \mathbb{R}_+^q$ is based on the definition of the production or attainable set

$$\Psi = \{(x, y) \in \mathbb{R}_+^p \times \mathbb{R}_+^q \mid x \text{ can produce } y\}. \tag{2.1}$$

The technology, or the efficient boundary, is defined as

$$\Psi^\partial = \{(x, y) \in \Psi \mid (\xi^{-1}x, \xi y) \notin \Psi \; \forall \, \xi > 1\}. \tag{2.2}$$

The "minimal" set of conditions usually imposed on Ψ (e.g., see Shephard, 1970, Färe, 1988) are included in the following assumption.

Assumption 2.1. Minimal Conditions
(i) Ψ is closed; (ii) the inputs and outputs are freely (or strongly) disposable, i.e., for all $(x, y) \in \Psi$, if $x' \geq x$ and $y' \leq y$, then $(x', y') \in \Psi$; and (iii) all production requires some non-zero inputs i.e., if $x = 0$, $y \geq 0$ and $y \neq 0$, then $(x, y) \notin \Psi$. [2]

Part (ii) of the assumption amounts to an assumption of weak monotonicity on the frontier. Part (iii) rules out free lunches by implying that something cannot be produced from nothing. Note that in this general representation, Ψ is not assumed to be convex. Similarly, there is no assumption on returns to scale. The technology exhibits constant returns to scale (CRS) if $(x, y) \in \Psi$ implies $(ax, ay) \in \Psi$ for all $a > 0$, but whether this is the case is ultimately an empirical question. We will see below that if we add these assumptions, when appropriates, to our models, we should adapt the estimators to obtain better precision. We will also see that these hypothesis can be tested.

[2] Here and below, inequalities on vectors are understood to be component-wise.

Once the production set as been defined, the efficiency of a production plan $(x, y) \in \Psi$ can be measured by its distance to the efficient boundary Ψ^{∂}. The most popular are the Debreu-Farrell (Debreu, 1951, Farrell, 1957) radial distances, i.e., the *radial input oriented* efficiency measure

$$\theta(x, y \mid \Psi) = \inf\{\theta \mid (\theta x, y) \in \Psi\} \leq 1 \tag{2.3}$$

and the *radial output oriented* measure

$$\lambda(x, y \mid \Psi) = \sup\{\lambda \mid (x, \lambda y) \in \Psi\} \geq 1. \tag{2.4}$$

Clearly, for a point $(x, y) \in \Psi$, $\theta(x, y \mid \Psi)$ represents the maximum, feasible proportionate reduction of input which is possible while producing output level y. Similarly, $\lambda(x, y \mid \Psi)$ represents the maximum, feasible proportionate increase of output quantities that is possible while using x levels of inputs.

More recently, alternative non-radial measures have been proposed. Färe et al. (1985) introduce the *hyperbolic* efficiency measure

$$\gamma(x, y \mid \Psi) = \inf\{\gamma \mid (\gamma x, \gamma^{-1} y) \in \Psi\} \leq 1, \tag{2.5}$$

which measures the feasible, simultaneous reduction in input quantities and increase in output quantities by the same proportion. Chambers et al. (1996, 1998) propose the additive, directional distance function

$$\beta(x, y; d_x, d_y \mid \Psi) = \sup\{\beta \mid (x - \beta d_x,\ y + \beta d_y) \in \Psi\} \geq 0 \tag{2.6}$$

where $d_x \in \mathbb{R}^p_+$ and $d_y \in \mathbb{R}^q_+$ are direction vectors with at least one element of d_x or d_y strictly greater than zero. The directional distance function measures distance to the boundary of the production set along the linear path given by by the direction vectors. It is easy to show that $\theta(x, y \mid \Psi) = 1 - \beta(x, y; x, 0_q \mid \Psi)$ and $\lambda(x, y \mid \Psi) = 1 + \beta(x, y; 0_p, y \mid \Psi)$, where 0_k denotes a k-vector of zeros.

Directional distance functions permit efficiency to be measured while holding some inputs or outputs constant. However, the additive nature of directional distance functions requires careful specification of the direction vectors d_x and d_y to avoid dependence on units of measurement. In particular, elements of the direction vectors should be defined in terms of the same units used to measure corresponding input and output levels. Wilson (2025) generalizes the hyperbolic measure in (2.5) to provide a *proportional,* multiplicative measure of efficiency while allowing some inputs or outputs to be held fixed. As discussed by Wilson (2025), empirical results obtained with the generalized hyperbolic distance function are easier to interpret than results based on the additive, directional measure in (2.6). See Wilson (2025) for details and discussion.

The cone spanned by Ψ with vertex at the origin given by

$$\mathcal{C}(\Psi) = \{(\tilde{x}, \tilde{y}) \mid \tilde{x} = ax,\ \tilde{y} = ay,\ \forall\, a \in \mathbb{R}_+ \text{ and } \forall\, (x, y) \in \Psi\} \tag{2.7}$$

is needed to define productivity indices. Similar to (2.2), the frontier of $\mathcal{C}(\Psi)$ is given by

$$\mathcal{C}^\partial(\Psi) = \{(x, y) \mid (x, y) \in \mathcal{C}(\Psi),\ (\xi^{-1}x, \xi y) \notin \mathcal{C}(\Psi)\ \forall\, \xi > 1\}. \tag{2.8}$$

Input, output, hyperbolic and directional distances from $(x, y) \in \mathbb{R}_+^{p+q}$ to the boundary of $\mathcal{C}(\Psi)$ are defined by replacing Ψ in (2.3)–(2.6) with $\mathcal{C}(\Psi)$. Moreover,

$$\theta(x, y \mid \mathcal{C}(\Psi))^{1/2} = \lambda(x, y \mid \mathcal{C}(\Psi))^{-1/2} = \gamma(x, y \mid \mathcal{C}(\Psi)) \tag{2.9}$$

by Kneip et al. (2021, Lemma 3.1) for the case where Ψ is convex and by Kneip et al. (2026, Lemma 2.1) when Ψ is not convex. In order to simplify the notation below, $\theta(x, y \mid \Psi)$ and $\theta(x, y \mid \mathcal{C}(\Psi))$ will be denoted by $\theta(x, y)$ and $\theta_C(x, y)$ when doing so does not create ambiguity.

It is important to note that since Ψ is not necessarily convex, $\mathcal{C}(\Psi)$ is not necessarily convex. In addition, if CRS prevails, then $\mathcal{C}(\Psi) = \Psi$ regardless of whether Ψ is convex. CRS does not require Ψ to be convex. In general, $\Psi \subseteq \mathcal{C}(\Psi)$. Since Ψ is unknown, $\mathcal{C}(\Psi)$ is unknown, and hence both must be estimated from a sample $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$. Convexity and CRS are strong assumptions, and whether either of these hold is ultimately an empirical question. Fortunately, today there are statistical results allowing researchers to test convexity versus non-convexity of Ψ as well as CRS versus non-CRS of the technology. The relevant tests are discussed below.

Remark 2.1. Convex or Non-Convex?
The production set Ψ is often assumed to be convex in the literature on efficiency analysis and in microeconomics textbooks. In some (perhaps many) situations, convexity may be a reasonable and appropriate assumption. But in some other cases, as confirmed in a number of empirical studies, the assumption is not supported by the data. See, for example, Apon et al. (2015), Daraio et al. (2021), O'Loughlin and Wilson (2021), Wilson (2021), Wilson and Zhao (2022) and Simar and Wilson (2024). Moreover, Koopmans (1951), Farrell (1957) and Afriat (1972) question whether convexity is an appropriate assumption. Kneip et al. (2026, Appendix A) includes a detailed discussion, with references to the literature, indicating that many authors have questioned the assumption of convexity based on economic theory but also

based on empirical observation in various industries. in the end, as noted in Afriat (1972), convexity or non-convexity of Ψ is an empirical issue that should be tested. Section 6 below discusses a procedure for testing convexity against non-convexity.

2.2 The DGP: A Statistical Model

2.2.1 Deterministic Frontiers

A model of the DGP is a set of assumptions specifying how data on inputs and outputs are generated. Here, the observed input-output pairs (X_i, Y_i) are assumed to be technically attainable and to be independently and identically distributed (iid). In other words, the input-output pairs are drawn randomly from a population of units distributed over the attainable set Ψ, implying

$$\Pr((X_i, Y_i) \in \Psi) = 1. \tag{2.10}$$

Models in the family of models incorporating this assumption are called *deterministic frontier models* in the efficiency literature, although this name is unfortunate and misleading because nothing is deterministic in (2.10).[3] By contrast, *Stochastic Frontier Analysis* (SFA) allows for the possibility of some limited noise in the DGP so that

$$\Pr((X_i, Y_i) \in \Psi) \leq 1. \tag{2.11}$$

Most of the non-parametric approaches to efficiency estimation rely on so-called deterministic models, since the SFA introduces additional identification issues. Models used with SFA are almost always parametric models incorporating restrictive assumptions, but allowing (large sample) identification. Recently, several non-parametric approaches for SFA have been proposed as discussed later in Section 8.

2.2.2 Regularity Assumptions

Depending on the economic assumptions on the model (e.g., convexity or non-convexity of Ψ, CRS or non-CRS of the frontier, etc.) and depending on the particular estimator to be used, the regularity assumptions are rather technical and may vary slightly depending on the estimator used to estimate efficiency (see Park et al., 2000, Kneip et al., 1998, Kneip et al., 2008, 2015b, 2021, 2026 and Park et al., 2010 for details). A summary is given in

[3] Wikipedia Contributors (2025) describes a "deterministic system" as a "system in which no randomness is involved in the development of future states of the system" and notes that "a deterministic model will thus always produce the same output from a given starting condition or initial state." This does not describe the assumption in (2.10).

Kneip et al. (2016, Appendix A.1). The main ideas are summarized below while avoiding technical details in the articles cited above as far as possible.

As noted above, Assumption 2.1 gives the minimal conditions typically imposed on Ψ. In addition, the discussion above leading to (2.10) is formalized in the following assumption.

Assumption 2.2. Random Sample
The sample (X_i, Y_i), $i = 1, \ldots, n$ are iid realizations of a random variable (X, Y) with probability density function $f(x, y)$ which has support $\Psi \subset \mathbb{R}_+^{p+q}$.

The joint density $f(x, y)$ implies the distribution function

$$H_{XY}(x, y) = \Pr(X \leq x, Y \geq y), \tag{2.12}$$

which gives the probability for a production plan (x, y) to be dominated. As noted by Cazals et al. (2002), under the free disposability assumption (Assumption 2.1(ii)) with Ψ either convex or non-convex, Ψ is the support of H_{XY} and so

$$\Psi = \{(x, y) \in \mathbb{R}_+^p \times \mathbb{R}_+^q \mid H_{XY}(x, y) > 0\}. \tag{2.13}$$

Assumption 2.3. Positiveness
The density $f(x, y)$ is strictly positive on the boundary Ψ^∂ and is continuous in any direction toward the interior of Ψ.

This assumption is common in frontier models (even in parametric models), but it can be relaxed as shown in Kneip et al. (1998) and Daouia et al. (2010, 2017) by allowing $f(x, y)$ to approach zero at the frontier, but at a cost of slower rates of convergence for efficiency estimators.

Finally, in order to prove most of the consistency results of the non-parametric estimators, sufficient smoothness of the production frontier is needed. Smoothness of the frontier can be characterized in terms of the distances defined above, as in the following assumption.

Assumption 2.4. Smoothness
For all (x, y) in the interior of Ψ, the functions $\theta(x, y|\Psi)$, $\lambda(x, y|\Psi)$, $\gamma(x, y|\Psi)$ and $\beta(x, y|\Psi)$ are sufficiently differentiable in both their arguments.

The required order of differentiability depends on the chosen estimator, as described in the references mentioned above.

3 Nonparametric Estimators

3.1 The Production Set

3.1.1 Basic Envelopment Estimators of Ψ

Deprins et al. (1984) propose the Free Disposal Hull (FDH) estimator of Ψ. This estimator requires free disposability of inputs and outputs, but neither convexity of Ψ nor CRS for the frontier. The estimator can be described in term of a non-parametric estimator of H_{XY}, i.e.,

$$\widehat{\Psi}_{FDH}(\mathcal{X}_n) = \{(x, y) \mid \widehat{H}_{XY}(x, y) > 0\}, \tag{3.1}$$

where $\widehat{H}_{XY}$ is the empirical version of H_{XY} given by

$$\widehat{H}_{XY}(x, y) = n^{-1} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x, Y_i \geq y). \tag{3.2}$$

It is easy to see that an equivalent definition is

$$\widehat{\Psi}_{FDH}(\mathcal{X}_n) = \left\{(x, y) \in \mathbb{R}_+^{p+q} \mid 0 \leq y \leq Y_i,\ x \geq X_i,\ (X_i, Y_i) \in \mathcal{X}_n\right\}, \tag{3.3}$$

which is simply the union of all the orthants in $\mathbb{R}_+^{p+q}$ lying southeast of the sample observations (X_i, Y_i), $i = 1, \ldots, n$.

Farrell (1957) suggests estimating Ψ by the convex hull of $\widehat{\Psi}_{FDH}$, and Charnes et al. (1978) described the corresponding input and output-oriented efficiency estimators in terms of linear programs, and Rhodes (1978) called this approach "Data Envelopment Analysis" (DEA). The DEA estimator of Ψ, i.e., the convex hull of $\widehat{\Psi}_{FDH}$, can be written as

$$\begin{aligned}\widehat{\Psi}_{DEA}(\mathcal{X}_n) = \Big\{(x, y) \in \mathbb{R}^{p+q} \mid y \leq \sum_{i=1}^{n} \xi_i Y_i,\ x \geq \sum_{i=1}^{n} \xi_i X_i, \\ \sum_{i=1}^{n} \xi_i = 1,\ \xi_i \geq 0 \ \forall\, i = 1, \ldots, n\Big\}.\end{aligned} \tag{3.4}$$

It is important to note that neither $\widehat{\Psi}_{FDH}$ nor $\widehat{\Psi}_{DEA}$ impose CRS on the technology. Either estimator can be used regardless of whether the technology is CRS, and the FDH estimator can be used regardless of whether Ψ is convex. But $\widehat{\Psi}_{DEA}$ is appropriate only if Ψ is convex.

Both $\widehat{\Psi}_{FDH}(\mathcal{X}_n)$ and $\widehat{\Psi}_{DEA}(\mathcal{X}_n)$ provide estimates of the boundary of the production set Ψ. However, the FDH estimator gives an estimate with a stair-step shape due to the union of the orthants in (3.3). Jeong and Simar (2006) propose a linearized FDH (LFDH) estimator $\widehat{\Psi}_{LFDH}(\mathcal{X}_n)$ that avoids the stair-step feature of the ordinary FDH estimator. The idea underlying the LFDH estimator is to link the corners of the orthants in the FDH estimator by linear interpolation. Consequently, $\widehat{\Psi}_{FDH}(\mathcal{X}_n) \subseteq \widehat{\Psi}_{LFDH}(\mathcal{X}_n)$.

3.1.2 Estimation of $\mathcal{C}(\Psi)$

As noted above, $\Psi \subseteq \mathcal{C}(\Psi)$ with $\Psi = \mathcal{C}(\Psi)$ if and only if CRS prevails. Under the free disposability assumption, the cone spanned by $\widehat{\Psi}_{FDH}$,

$$\widehat{\Psi}_{CFDH}(\mathcal{X}_n) = \big\{(x, y) \in \mathbb{R}_+^{p+q} \mid (x, y) = (a\widetilde{x}, a\widetilde{y}) \text{ for some } a \geq 0 \text{ and } (\widetilde{x}, \widetilde{y}) \in \widehat{\Psi}_{FDH}(\mathcal{X}_n)\big\}, \tag{3.5}$$

serves an an estimator of $\mathcal{C}(\Psi)$. If, in addition, CRS holds then then $\widehat{\Psi}_{CFDH}(\mathcal{X}_n)$ also estimates $\Psi = \mathcal{C}(\Psi)$.

In cases where both free disposability and convexity hold for Ψ, then the conical hull of $\widehat{\Psi}_{DEA}(\mathcal{X}_n)$ given by

$$\widehat{\Psi}_{CDEA}(\mathcal{X}_n) = \left\{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq \sum_{i=1}^{n} \xi_i Y_i;\ x \geq \sum_{i=1}^{n} \xi_i X_i,\ \xi_i \geq 0 \ \forall\, i = 1,\ \ldots,\ n\right\} \tag{3.6}$$

is the natural estimator of $\mathcal{C}(\Psi)$. The estimator in (3.6) is similar to the one in (3.4), but the equality constraint on the multipliers ξ has been dropped in (3.6). Note also that under both (i) free disposability and (ii) convexity of Ψ one could also use $\widehat{\Psi}_{CFDH}(\mathcal{X}_n)$ as an estimator of $\mathcal{C}(\Psi)$, but as seen below ignoring convexity when it holds results in a loss of precision in the resulting estimates of the efficiency scores.

If the CRS assumption holds in addition to both free disposability and convexity of Ψ, then both $\widehat{\Psi}_{CFDH}(\mathcal{X}_n)$ and $\widehat{\Psi}_{CDEA}(\mathcal{X}_n)$ estimate Ψ, but $\widehat{\Psi}_{CDEA}(\mathcal{X}_n)$ results in estimates with better precision. But, it is important to note that *if CRS does not hold*, then neither $\widehat{\Psi}_{CDEA}(\mathcal{X}_n)$ (which also require that Ψ is convex) nor $\widehat{\Psi}_{CFDH}(\mathcal{X}_n)$ (regardless of whether Ψ is convex) estimate $\mathcal{C}(\Psi) \neq \Psi$. Nonetheless, the conical estimators $\widehat{\Psi}_{CFDH}(\mathcal{X}_n)$ and $\widehat{\Psi}_{CDEA}(\mathcal{X}_n)$ are needed for estimating MPIs, even when CRS does not hold.

3.2 Efficiency Scores

Section 2 shows how the various efficiency scores are defined in terms of the unknown Ψ. Nonparametric estimators are obtained by "plugging in" an appropriate non-parametric estimator for Ψ, i.e., by replacing Ψ in (2.3)–(2.6) with an appropriate estimator of Ψ. The appropriateness of an estimator of Ψ depends on what assumptions about Ψ hold. Below, in Section 3.2.1, a detailed presentation is given for the input-oriented case, and Section 3.2.2 summarizes the main expressions for other orientations.

3.2.1 Input-Oriented Case

For general technologies where only free disposability is assumed for Ψ, the FDH estimator of $\theta(x, y \mid \Psi)$ is obtained by replacing Ψ in (2.3) with $\widehat{\Psi}_{FDH}(\mathcal{X}_n)$ to obtain

$$\widehat{\theta}_{FDH}(x, y) = \inf \left\{ \theta \mid (\theta x, y) \in \widehat{\Psi}_{FDH}(\mathcal{X}_n) \right\}. \tag{3.7}$$

Some simple algebra (see Deprins et al., 1984) leads to

$$\widehat{\theta}_{FDH}(x, y) = \min_{i \in D(x,y)} \left[\max_{k=1,\dots,p} \left(\frac{X_{ik}}{x_k} \right) \right], \tag{3.8}$$

where $D(x, y) = \{i \mid X_i \leq x,\ Y_i \geq y\}$ is the set of data points dominating (x, y). Similarly, the distance to the boundary of $\mathcal{C}(\Psi)$ is estimated by

$$\widehat{\theta}_{CFDH}(x, y) = \inf \left\{ \theta \mid (\theta x, y) \in \widehat{\Psi}_{CFDH}(\mathcal{X}_n) \right\}.$$

Kneip et al. (2026) show that this is equivalent to

$$\widehat{\theta}_{CFDH}(x, y) = \min_{i=1,\, \dots,\, n} \left[\frac{\max_{k=1,\dots,p} \left(\frac{X_{ik}}{x_k} \right)}{\min_{j=1,\, \dots,\, q} \left(\frac{Y_{ij}}{y_j} \right)} \right]. \tag{3.9}$$

Of course, $\widehat{\theta}_{CFDH}(x, y)$ estimates the efficiency of the production plan (x, y) under the CRS assumption for Ψ without imposing convexity as discussed earlier.

When one is willing to assume Ψ is convex, DEA estimators can be used. The DEA estimator of $\theta(x, y \mid \Psi)$ is

$$\widehat{\theta}_{DEA}(x, y) = \inf \left\{ \theta \mid (\theta x, y) \in \widehat{\Psi}_{DEA}(\mathcal{X}_n) \right\}, \tag{3.10}$$

which can be computed by solving the linear program

$$\widehat{\theta}_{DEA}(x, y) = \inf_{\theta, \xi} \left\{ \theta \mid y \leq \sum_{i=1}^{n} \xi_i Y_i;\ \theta x \geq \sum_{i=1}^{n} \xi_i X_i,\ \sum_{i=1}^{n} \xi_i = 1;\ \xi_i \geq 0\ \forall\ i = 1, \dots, n \right\} \tag{3.11}$$

as suggested by Banker et al. (1984). The estimator in (3.10)–(3.11) is often referred to in the literature as the variable-returns-to-scale (VRS) DEA (VRS-DEA) estimator since it allows varying (i.e., increasing, followed by locally constant and then decreasing) returns to scale for the frontier of Ψ.

Similarly, distance to the boundary of $\mathcal{C}(\Psi)$ is estimated by

$$\begin{aligned}\widehat{\theta}_{CDEA}(x,y) &= \inf\left\{\theta \mid (\theta x, y) \in \widehat{\Psi}_{CDEA}(\mathcal{X}_n)\right\} \\ &= \inf_{\theta,\xi}\left\{\theta \mid y \leq \sum_{i=1}^{n}\xi_i Y_i;\ \theta x \geq \sum_{i=1}^{n}\xi_i X_i;\ \xi_i \geq 0\ \forall\ i = 1,\ldots,n\right\} \end{aligned} \quad (3.12)$$

provided Ψ is convex. If both convexity of Ψ and CRS hold, then $\widehat{\theta}_{CDEA}(x,y)$ estimates the efficiency of the production plan (x,y). In such cases the estimator is often referred to as the CRS-DEA estimator. Otherwise, the estimator in (3.12) remains useful for estimating MPIs.

3.2.2 Other Orientations

Expressions similar to those in Section 3.2.1 are available for the output-oriented case. In particular, the output-oriented analogs of (3.8) and (3.9) are (respectively)

$$\widehat{\lambda}_{FDH}(x,y) = \max_{i \in D(x,y)}\left[\min_{j=1,\ldots,q}\left(\frac{Y_{ij}}{y_j}\right)\right] \quad (3.13)$$

and

$$\widehat{\lambda}_{CFDH}(x,y) = \max_{i=1,\ldots,n}\left[\frac{\min_{j=1,\ldots,q}\left(\frac{Y_{ij}}{y_j}\right)}{\max_{k=1,\ldots,p}\left(\frac{X_{ik}}{x_k}\right)}\right]. \quad (3.14)$$

Note that Kneip et al. (2026) show that $\widehat{\lambda}_{CFDH}(x,y) = \left[\widehat{\theta}_{CFDH}(x,y)\right]^{-1}$.

The output-oriented, DEA analogs of (3.10)–(3.11) are given by

$$\widehat{\lambda}_{DEA}(x,y) = \sup_{\lambda,\xi}\left\{\lambda \mid \lambda y \leq \sum_{i=1}^{n}\xi_i Y_i;\ x \geq \sum_{i=1}^{n}\xi_i X_i;\ \sum_{i=1}^{n}\xi_i = 1;\ \xi_i \geq 0\ \forall\ i = 1,\ \ldots,;n\right\}. \quad (3.15)$$

and

$$\widehat{\lambda}_{CDEA}(x,y) = \sup_{\lambda,\xi}\left\{\lambda \mid \lambda y \leq \sum_{i=1}^{n}\xi_i Y_i;\ x \geq \sum_{i=1}^{n}\xi_i X_i;\ \xi_i \geq 0\ \forall\ i = 1,\ldots,n\right\}, \quad (3.16)$$

respectively. The estimator in (3.16) can be used as an estimator of efficiency if both convexity and CRS hold. If only convexity of Ψ holds but CRS does not hold, then $\widehat{\lambda}_{CDEA}(x,y)$ remains useful for estimating MPIs.

Wilson (2011) provides estimators of the hyperbolic efficiency measure defined in (2.5). In particular, the FDH estimator of $\gamma(x, y \mid \Psi)$ is given by

$$\widehat{\gamma}_{FDH}(x,y) = \min_{i=1,\,\ldots,\,n} \left[\max_{\substack{k\,=\,1\,\ldots,\,p \\ j\,=\,1,\ldots,\,q}} \left(\frac{X_{ik}}{x_k},\ \frac{y_j}{Y_{ij}} \right) \right], \tag{3.17}$$

which requires neither convexity of Ψ nor CRS. By Kneip et al. (2026, Lemma 3.1), (2.9),

$$\widehat{\gamma}_{CFDH}(x,y) = \left[\widehat{\theta}_{CFDH}(x,y)\right]^{1/2} = \left[\widehat{\lambda}_{CFDH}(x,y)\right]^{-1/2}. \tag{3.18}$$

The estimator $\widehat{\gamma}_{CFDH}(x,y)$ can be used to estimate efficiency when CRS holds, and is useful for estimating MPIs regardless of whether CRS holds.

The DEA estimator of $\gamma(x, y \mid \Psi)$ is given by

$$\widehat{\gamma}_{DEA}(x,y) = \inf_{\gamma,\xi} \left\{ \gamma \mid \gamma^{-1} y \leq \sum_{i=1}^{n} \xi_i Y_i;\ \gamma x \geq \sum_{i=1}^{n} \xi_i X_i;\ \sum_{i=1}^{n} \xi_i = 1;\ \xi_i \geq 0\ \forall\ i = 1,\ \ldots,\ n \right\} \tag{3.19}$$

and is appropriate for efficiency estimation when Ψ is convex. Computing (3.19) directly is a formidable challenge, but Wilson (2011) provides a numerical algorithm allowing computation of $\widehat{\gamma}_{DEA}(x,y)$ to an arbitrary degree of accuracy. By Kneip et al. (2021, Lemma 3.2),

$$\widehat{\gamma}_{CDEA}(x,y) = \left[\widehat{\theta}_{CDEA}(x,y)\right]^{1/2} = \left[\widehat{\lambda}_{CDEA}(x,y)\right]^{-1/2}. \tag{3.20}$$

The estimator $\widehat{\gamma}_{CDEA}(x,y)$ can be used to estimate efficiency when Ψ is convex and CRS holds, and is useful for estimating MPIs regardless of whether CRS holds provided Ψ is convex. Wilson (2025) provides both FDH and DEA estimators for the generalized hyperbolic distance function where some inputs or outputs are held constant.

Simar and Vanhems (2012), Simar et al. (2012), Daraio et al. (2020) and Daraio et al. (2025) develop estimators for additive, directional efficiency measures. distances case. The presentation here is limited to the case where $d_x > 0$ and $d_y > 0$, i.e., where all the elements of the distance vectors are strictly positive. The cases where some elements are equal to zero complicate the notation and are described in detail in the references given above.

The FDH estimator of $\beta(x, y; d_x, d_y \mid \Psi)$ defined in (2.6) is given by

$$\widehat{\beta}_{FDH}(x,y;d_x,d_y) = \max_{i=1,\ldots,n} \left[\min_{\substack{k\,=\,1\,\ldots,\,p \\ j\,=\,1,\ldots,\,q}} \left\{ x_k^{\star} - X_{ik}^{\star}, Y_{ij}^{\star} - y_j^{\star} \right\} \right], \tag{3.21}$$

where $X^\star = X \oslash d_x$, $x^\star = x \oslash d_x$, $Y^\star = Y \oslash d_y$, $y^\star = y \oslash d_y$, and $\oslash$ denotes the Hadamard component-wise division of vectors.

Computation of the directional distance to the boundary of the cone spanned by Ψ is more complicated. In the general case, Daraio et al. (2025) show that

$$\widehat{\beta}_{CFDH}(x,y;d_x,d_y) = \sup_{a>0}\left\{\frac{1}{a}\max_{i=1,\ldots,n}\left\{\min_{\substack{k=1,\ldots,p\\ j=1,\ldots,q}}\left[ax_k^\star - X_{i,k}^\star, Y_{i,j}^\star - ay_j^\star\right]\right\}\right\}, \tag{3.22}$$

which involves a univariate optimization problem. Daraio et al. (2025) provide an easy algorithm to solve this numerical problem and show how the computations simplify when $d_x = x$ and $d_y = y$.

Simar et al. (2012) give the DEA estimator

$$\begin{aligned}\widehat{\beta}_{DEA}(x,y;d_x,d_y) = \sup_{\beta,\xi}\Big\{\beta \mid y+\beta d_y \leq \sum_{i=1}^n \xi_i Y_i;\ x-\beta d_x \geq \sum_{i=1}^n \xi_i X_i \\ \sum_{i=1}^n \xi_i = 1;\ \xi_i \geq 0\ \forall\ i=1,\ldots,n\Big\}\end{aligned} \tag{3.23}$$

for cases where Ψ is convex. Finally, the DEA, additive, directional estimator of distance to the boundary of the cone spanned by Ψ is denoted by $\widehat{\beta}_{CDEA}(x,y;d_x,d_y)$ and is given by the solution to (3.23) after dropping the constraint $\sum_{i=1}^n \xi_i = 1$ on the multipliers in (3.23). Provided Ψ is convex and CRS holds, $\widehat{\beta}_{CDEA}(x,y;d_x,d_y)$ estimates the efficiency of a unit (x,y) in the chosen direction.

3.3 Statistical Properties and Inference

Although Farrell (1957) and Deprins et al. (1984) make the first applications of DEA and FDH estimators in empirical exercises, the first to view the results of FDH and DEA methods as *estimates* are Simar (1992, 1996), Banker (1993) and Korostelev et al. (1995a, 1995b). The papers by Korostelev et al. are the first to derive, in a well-defined statistical model, the rates of convergence of the FDH and DEA estimators and to demonstrate some optimal properties within some larger classes of estimators. As shown below, these results have been extended to various cases, providing the asymptotic properties needed to make inference. The following discussion again focuses on the input, radial case to avoid excessive notation and complication. All of the the properties discussed below extend to each of the directions described above with appropriate adaptations of the regularity conditions.

3.3.1 Individual Efficiencies

Under appropriate assumptions, all of the efficiency estimators discussed in Sections 3.2.1–3.2.2 provide statistically consistent estimates of efficiency for a producer operating at a point $(x, y) \in \Psi$. In addition, the sampling distribution of each estimator discussed above converges to a non-degenerate limiting distribution at rate n^κ as $n \to \infty$. The rate of convergence gives an idea, in a probabilistic sense, of the order of estimation error. For the input orientation, convergence rate n^κ and a fixed point (x, y), $\widehat{\theta}_\bullet(x, y) - \theta_\bullet(x, y) = O_p\left(n^{-\kappa}\right)$, where "$\bullet$" represents either "FDH," "CFDH," "DEA" or "CDEA". Similar statements hold for the output, hyperbolic, generalized hyperbolic, and directional measures and their estimators.

Table 1 gives the values of κ for each of the FDH, CFDH, DEA and CDEA estimators depending on whether Ψ is convex and whether the frontier is CRS (free disposability is required in each case). In most cases, the rates in Table 1 are slower than the usual rate of $n^{1/2}$ that obtains with parametric estimators in correctly-specified models. Also, in each case in Table 1 the convergence rate becomes slower as dimensionality given by $(p + q)$ increases, reflecting the well-known *curse of dimensionality* that is common among non-parametric estimators.

Table 1: Values of κ for the Various Efficiency Estimators

Assumptions			**Estimators**			
FD	CONV	CRS	FDH	CFDH	DEA	CDEA
Y	N	N	$1/(p+q)$	$1/(p+q-0.5)$	—	—
Y	N	Y	$1/(p+q)$	$1/(p+q-1)$	—	—
Y	Y	N	$1/(p+q)$	$1/(p+q-0.5)$	$2/(p+q+1)$	$2/(p+q+1)$
Y	Y	Y	$1/(p+q)$	$1/(p+q-1)$	$2/(p+q)$	$2/(p+q)$

NOTE: "Y" indicates assumption is satisfied while "N" indicates assumption is not satisfied. "FD" and "CONV" indicate "free disposability" and "convexity" (respectively). FDH and CFDH estimators are consistent in each case, but DEA and CDEA estimators are not consistent when convexity is not satisfied.

Under the appropriate, mild regularity conditions,

$$n^\kappa \left(\widehat{\theta}_\bullet(x, y) - \theta_\bullet(x, y)\right) \xrightarrow{\mathcal{L}} Q_\bullet(\eta(x, y)) \tag{3.24}$$

$n \to \infty$, where $Q_\bullet(\eta)$ is some regular non-degenerate distribution depending on a vector $\eta(x, y)$ of unknowns and the particular estimator, and where again "$\bullet$" represents either "FDH," "CFDH," "DEA" or "CDEA". The values of κ shown in Table 1 reveal that when

CRS holds, CFDH, DEA and CDEA estimators have faster rates than when CRS does not hold. The FDH and CFDH rates are unaffected by whether Ψ is convex. If Ψ is convex then the DEA and CDEA estimators can be used, providing faster rates of convergence than the FDH and CFDH estimators (respectively). The rate of the ordinary FDH estimator is not affected by whether Ψ is convex nor by whether CRS holds, and is slower than the rates of the CFDH, DEA and CDEA estimators for the same number $(p+q)$ of dimensions. At the same time, the ordinary FDH estimator is consistent regardless of whether convexity or CRS hold. Consequently, it is important to test whether convexity of Ψ holds and whether CRS holds. Tests for both are discussed below in Section 6.

Under mild assumptions including free disposability assumption for Ψ (but not convexity nor CRS), the limiting distribution $Q_\bullet(\eta(x,y))$ for the FDH and CFDH estimators is a Weibull distribution. For example, for the FDH estimator,

$$n^{1/(p+q)}\left(\widehat{\theta}_{FDH}(x,y)-\theta(x,y)\right) \xrightarrow{\mathcal{L}} \text{Weibull}(\eta(x,y)) \tag{3.25}$$

as $n\to\infty$, where $\eta(x,y)$ is a complicated function that depends on the DGP, p, q and (x,y) as proved by Park et al. (2000). Kneip et al. (2026) derive a similar result for the CFDH estimator.

The LFDH (i.e., linearized FDH) estimator mentioned above can also be used under the same assumptions required for the FDH estimator. Jeong and Simar (2006) show that both estimators share similar asymptotic properties. However, the Monte Carlo results of Jeong and Simar reveal that efficiency estimates obtained with the LFDH estimator have better bias and variance in finite samples than estimates obtained with the FDH estimator.

Gijbels et al. (1999) are the first to derive the asymptotic distribution of the basic DEA estimator under appropriate mild conditions including both free disposability and convexity of Ψ, but only for the case where $p=q=1$. Jeong and Park (2006) extend the latter to case of $p=1$ and $q>1$ for the input oriented case. Kneip et al. (1998) and Kneip et al. (2008, 2015b, 2021) show in the general multivariate case, that the limiting distribution $Q_\bullet(\eta)$ for the DEA efficiency estimator under non-CRS is a well-defined, non-degenerate distribution, but no closed form representation exists. The CRS-convex case is also treated by Park et al. (2010) establish asymptotic properties of the CDEA estimator under CRS, and suggest a method for simulating the limiting distribution using smoothing techniques. Kneip et al. (2016) develop the asymptotic properties of the DEA estimator under CRS, and prove that the estimator attains the faster rate of the CDEA estimator under CRS, while Kneip et al. (2021) establish that under non-CRS, the CDEA estimator has the slower rate of the ordinary DEA estimator.

The results above are presented in terms of the input-oriented estimators, but the results extend obviously to the output-oriented case with changes in notation. The results are extended to the hyperbolic measure by Wilson (2011) and to the generalized hyperbolic measure by Wilson (2025), and to directional distances by Simar and Vanhems (2012) and Simar et al. (2012). In all cases, the results are analogous to those discussed above, and the convergence rates are as shown in Table 1.

Remark 3.1. Dimension reduction

Clearly, as the number $(p+q)$ *of dimensions increases, the order of the estimation error increases for a given sample size* n *since the rates given by* n^{κ} *become much slower than the usual parametric rate* $n^{1/2}$*. Moreover, the ability to make accurate inference decreases as dimensionality grows since inference depends on asymptotic approximations for finite-sample distributions of the estimators. Table 2 shows the magnitudes of the five expressions for* κ *for* $(p+q) \in \{2,\ 3,\ \ldots,\ 15\}$*. The results in Table 2 confirm that the FDH estimator achieves the parametric, root-n rate only when* $(p+q) = 2$*. None of the estimators that have been discussed achieve the parametric rate when* $(p+q) > 4$*. When the number of dimensions is 12 or more, which can be found in the literature, the values of* κ *(and rates of convergence) are too small (slow) to give meaningful results even with hundreds of observations.*

Wilson (2018) introduces the notion of effective parametric sample size *as the sample size* m *resulting in an estimation with the same order of error than one would achieve in a parametric model. It is easy to see that* $m \approx n^{2\kappa}$*, which can be very small if* $(p+q)$ *is large. For example, if the FDH estimator is used with* $n = 1000$ *and* $p+q = 5$*, then* $m \approx 16$*, and one should expect substantial estimation error and imprecise inference.*

Dimension reduction can be helpful. In particular, in production data, inputs as well as outputs are often highly correlated, and hence can be summarized by an appropriate linear combination. Daraio and Simar (2007a) suggest using eigenvalue decompositions of the second moment matrices of the inputs (or outputs) to derive these linear combinations. Wilson (2018) provides a number of diagnostic tools to check the appropriateness of dimension reduction, as well as illustrate in several famous examples from the efficiency literature. Wilson (2018) shows results from a comprehensive set of Monte Carlo experiments indicating when dimension-reduction is likely to be useful as well as how much is likely to be gained in terms of reduced estimation error when dimension-reduction is employed.

Remark 3.2. Parametric rate versus non-parametric rate

Comparisons of convergence rates in non-parametric problems with the parametric root-n rate is often specious. Although non-parametric rates are often slower than the parametric *rate, the parametric rate is only relevant when the statistical model is correctly specified. Otherwise, estimates in a wrongly-specified parametric model converge rapidly to some feature that is irrelevant to the DGP that produced the researcher's data. Robinson (1988) calls such estimators "root-n inconsistent" as opposed to root-n consistent." If one views the space of possible DGPs as continuous, and hence uncountable, then it is hard to escape the conclusion that a parametric model chosen for its ease of estimation (e.g., a translog production function) is almost surely a mis-specified model. Of course, in some cases a parametric model may provide a useful approximation to the underlying DGP, but this can only be known after careful testing, which is often not done.*

Table 2: Values of κ for Various Numbers of Dimensions and Efficiency Estimators

	κ				
$(p+q)$	$1/(p+q)$	$1/(p+q-0.5)$	$1/(p+q-1)$	$2/(p+q+1)$	$2/(p+q)$
2	0.5000	0.6667	1.0000	0.6667	1.0000
3	0.3333	0.4000	0.5000	0.5000	0.6667
4	0.2500	0.2857	0.3333	0.4000	0.5000
5	0.2000	0.2222	0.2500	0.3333	0.4000
6	0.1667	0.1818	0.2000	0.2857	0.3333
7	0.1429	0.1538	0.1667	0.2500	0.2857
8	0.1250	0.1333	0.1429	0.2222	0.2500
9	0.1111	0.1176	0.1250	0.2000	0.2222
10	0.1000	0.1053	0.1111	0.1818	0.2000
11	0.0909	0.0952	0.1000	0.1667	0.1818
12	0.0833	0.0870	0.0909	0.1538	0.1667
13	0.0769	0.0800	0.0833	0.1429	0.1538
14	0.0714	0.0741	0.0769	0.1333	0.1429
15	0.0667	0.0690	0.0714	0.1250	0.1333

3.3.2 The Bootstrap

The results on existence of non-degenerate limiting distributions, rates of convergence, etc. discussed above are necessary for statistical inference. However, the theoretical results, while needed, are not enough to be able to make inference in empirical applications due to the fact that either the form of the limiting distribution is unknown, or if it is known it involves unknown parameters that are difficult to estimate. In many cases, bootstrap methods, when correctly applied, can be used to make inferences.

The first suggestion of using the bootstrap for making inference in productivity and efficiency analysis appears in Simar (1992) for a panel-data setting, but no theoretical justification is given. Simar and Wilson (1998) is the first study suggesting bootstrap techniques for approximating the limiting distribution in the case of the basic DEA estimator. The idea seems rather simple: a random bootstrap sample $\mathcal{X}_n^*$ is generated from the original sample $\mathcal{X}_n$ in a way to mimic the DGP that generates the original sample $\mathcal{X}_n$. This leads to an estimator $\widehat{\Psi}_{DEA}(\mathcal{X}_n^*)$ which may be viewed as an estimator of $\widehat{\Psi}_{DEA}(\mathcal{X}_n)$, which in turn leads to bootstrap estimates $\widehat{\theta}^*_{DEA}(x, y)$ of the original efficiency estimates $\widehat{\theta}_{DEA}(x, y)$ obtained using the original sample. If the bootstrap is consistent, then

$$n^{2/(p+q+1)}\left(\widehat{\theta}^*_{DEA}(x, y) - \widehat{\theta}_{DEA}(x, y)\right) \overset{\text{approx.}}{\sim} n^{2/(p+q+1)}\left(\widehat{\theta}_{DEA}(x, y) - \theta(x, y)\right), \tag{3.26}$$

with the approximation improving as $n \to \infty$. Moreover, the distribution on the left-hand side of (3.26) can be approximated by Monte-Carlo replications, thereby providing a useful approximation to the unknown distribution on the right-hand side.

The main problem is to generate $\mathcal{X}_n^*$ in a way that ensures (3.26) holds. It is well-known that the naive bootstrap (i.e., resampling with replacement from the data in $\mathcal{X}_n$) is not consistent due to the unknown boundary of Ψ. Simar and Wilson (1998) propose using a smooth bootstrap, under the simplest model where the inefficiencies have the same distribution along the chosen direction (e.g., input rays) in the input-output space. Simar and Wilson (2000) extend the idea to more general settings to allow for heterogeneity in the distribution of efficiency. Simar and Wilson (1998, 2000) provide simulation evidence suggesting that the proposed methods work well in applied settings, but do not formally prove consistency of their bootstrap methods. Kneip et al. (2008) develop the full theory on the asymptotic properties of the DEA estimator and provide two consistent bootstrap algorithms. The first bootstrap requires smoothing not only the distribution of efficiency, but also the initial frontier estimate. This involves formidable computational challenges. A simplified version of this method is given by Kneip et al. (2011).

The second bootstrap of Kneip et al. (2008) involves sub-sampling, where a bootstrap sample $\mathcal{X}_m^*$ of size $m = n^\delta$ for some $\delta \in (0, 1)$ is drawn from the empirical distribution of the original sample $\mathcal{X}_n$ (Jeong and Simar, 2006 also prove that the sub-sampling bootstrap provides consistent inference when FDH estimators are used). But, the simulation results of Kneip et al. (2008) indicate that the choice of δ, which determines the size m of the sub-samples, is critical for reliable inference in finite samples. Simar and Wilson (2011a) provide a data-based algorithm based on work by Politis et al. (2001) for selecting an appropriate value of m in practice, for both FDH and DEA cases. Monte Carlo evidence provided by Simar and Wilson (2011a) shows that the method works well, even for moderate sample sizes. The optimal value of m is often well less than $n/2$, enhancing the computational speed of the procedure.

3.3.3 Central Limit Theorems

The bootstrap methods described above in Section 3.3.2 are useful for estimating confidence intervals for the efficiency of individual producers, evaluating the bias and variance of efficiency estimators, etc. Testing hypotheses about the structure of the underlying DGP (e.g., convexity versus non-convexity of Ψ, CRS versus non-CRS, whether the frontier is the same across different groups of producers, etc.) requires more work. In testing problems, one tests a null hypothesis denoted by H_0 against an alternative denoted by H_1. The null hypothesis is an assumption that may or may not hold; the goal is to determine whether the assumption is supported by the observed data. By construction, a statistical tests either (i) rejects H_0 (in which case H_1 is accepted, since the data provide evidence against H_0), or (ii) fails to reject H_0. Failure to reject the null does not imply that the null hypothesis is true, however. There are many reasons why a test might fail to reject the null even when it is false. For example, one may have too few data. Or, the null might be false, but not grossly so, and the available data are too few to indicate with a reasonable degree of confidence that the null is false.

In the usual framework, one defines a statistic $T(\mathcal{X}_n)$ which behaves differently according to whether the null hypothesis holds, thereby allowing the researcher to discriminate between the two assumptions. In the context of efficiency analysis, some early studies including Simar and Wilson (2001, 2002, 2011a) and Simar and Zelenyuk (2006, 2007) attempt to adapt the bootstrap methods described above to various testing problems. In each case, the test statistics involve FDH, DEA or CDEA estimators, depending on the hypothesis to be tested. These studies provide Monte Carlo evidence that seems to suggest that the proposed procedures achieve reasonable size and power in moderate sample sizes. However, the studies listed above lack theoretical results and give no convincing theoretical justification. This deficiency is remedied by Kneip et al. (2015b).

For illustration purposes, let $\widehat{\theta}(x, y \mid \mathcal{X}_n)$ denote the appropriate chosen estimator (FDH, DEA, CFDH or, CDEA) of the unknown efficiency $\theta(x, y)$ at a fixed point (x, y), where the notation in the estimator makes explicit that the sample $\mathcal{X}_n$ for estimation. In most cases, a test statistic $T(\mathcal{X}_n)$ will be a continuous function of the sample mean

$$\widehat{\mu}_n = n^{-1} \sum_{i=1}^{n} \widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) \tag{3.27}$$

of the efficiency estimator evaluated at the random points $(X_i, Y_i) \in \mathcal{X}_n$. Clearly, this is an estimator of population mean $\mu_\theta = \mathbb{E}(\theta(X, Y))$. If the true values of $\theta(X, Y)$ were known, one could use

$$\overline{\theta}_n = n^{-1} \sum_{i=1}^{n} \theta(X_i, Y_i) \tag{3.28}$$

for making inference about μ_θ because under mild regularity conditions $\sqrt{n}(\overline{\theta}_n - \mu_\theta) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2)$ by the Lindeberg-Levy CLT or Lindeberg-Feller CLT, where $\sigma_\theta^2 = \mathbb{V}(\theta(X, Y))$ is the population variance of the efficiency scores. But of course $\overline{\theta}_n$ is not observable, and one must use $\widehat{\mu}_n$ which is observable. Kneip et al. (2015b, Theorems 4.1–4.4) provide CLTs for μ_θ using $\widehat{\mu}_n$. The CLT results are based on properties of the moments of the FDH and DEA efficiency estimators and the CDEA estimator under CRS established by Kneip et al. (2015b, Theorems 3.1–3.3). These results are extended to the CDEA estimator under non-CRS by Kneip et al. (2021) and to the CFDH estimator by Kneip et al. (2026). The various results on moments of the estimators are summarized by the following.

Theorem 3.1. *Under appropriate regularity assumptions, as* $n \to \infty$

$$E\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right) = Cn^{-\kappa} + R_{n,\kappa}, \tag{3.29}$$

where $R_{n,k} = o\left(n^{-\kappa}\right)$,

$$E\left(\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i)\right)^2\right) = o\left(n^{-\kappa}\right), \tag{3.30}$$

and

$$\left| COV\left(\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \theta(X_i, Y_i),\ \widehat{\theta}(X_j, Y_j \mid \mathcal{X}_n) - \theta(X_j, Y_j)\right) \right| = o\left(n^{-1}\right) \tag{3.31}$$

for all $i, j \in \{1, \ldots, n\}$, $i \neq j$.

The values of the constant C, the rate κ, and the remainder term $R_{n,\kappa}$ depend on which estimator is used. The values of κ are given in Table 1 above for each estimator. Of course, the results outlined here depend on the corresponding relevant assumptions for individual efficiency scores as discussed above. See Kneip et al., 2015b, 2021, 2026 for details and discussion.

As shown by Kneip et al. (2015b, 2021 and 2026), the results in Theorem 3.1 lead to CLTs for μ_θ along the lines of the following theorem.

Theorem 3.2. *Under appropriate regularity conditions, as $n \to \infty$,*

(i) $\widehat{\mu}_n = n^{-1} \sum_{i=1}^{n} \widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n)$ *converges in probability to μ_θ with bias of order $O\big(n^{-\kappa}\big)$;*

(ii) $\sqrt{n}\,(\widehat{\mu}_n - \mu_\theta - Cn^{-\kappa} - R_{n,\kappa}) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2)$ *for some $C > 0$, and $R_{n,\kappa} = o(n^{-\kappa})$; and*

(iii) $\widehat{\sigma}_\theta^2 = n^{-1} \sum_{i=1}^{n} (\widehat{\theta}(X_i, Y_i \mid \mathcal{X}_n) - \widehat{\mu}_n)^2$ *is a consistent estimator of σ_θ^2.*

Clearly, the bias term $Cn^{-\kappa} + R_{n,\kappa}$ vanishes when multiplied by $\sqrt{n}$ if and only if $\kappa > 1/2$. The bias is constant when $\kappa = 1/2$, and the bias explodes to infinity whenever $\kappa < 1/2$ (Table 2 shows that this happens when $p + q > 2$, 3 or 4 depending on the particular case). Kneip et al. (2015b, 2016, 2021 and 2026) solve this problem by providing an estimator of the leading term $Cn^{-\kappa}$ of the bias, which is enough if $\kappa \geq 1/2$. But, more is needed for eliminating the remainder $R_{n,\kappa} = o(n^{-\kappa})$ when $\kappa < 1/2$. This is achieved by using sample means of random sub-samples of the estimated efficiencies. The main ideas are summarized below.

The leading term $Cn^{-\kappa}$ of the bias can be estimated using a generalized jackknife estimator. The sample $\mathcal{X}_n$ is split randomly into two independent, mutually exclusive, collectively exhaustive parts $\mathcal{X}_{n/2}^{(j)}$, $j \in \{1,\ 2\}$ of size $n/2$ (here, n is assumed even to simplify notation, but whether n is even or odd makes no difference asymptotically). Sample means of the efficiency estimates using only the sample $\mathcal{X}_{n/2}^{(j)}$ are computed, leading to the means

$$\widehat{\mu}_{n/2}^{(j)} = 2n^{-1} \sum_{i \mid (X_i, Y_i) \in \mathcal{X}_{n/2}^{(j)}} \widehat{\theta}(X_i, Y_i) \mid \mathcal{X}_{n/2}^{(j)}) \tag{3.32}$$

for $j \in \{1,\ 2\}$. The sample means $\widehat{\mu}_{n/2}^{(1)}$ and $\widehat{\mu}_{n/2}^{(2)}$ given by (3.32) each provide consistent estimates of μ_θ based on $n/2$ observations, so that the leading bias term is now $C(n/2)^{-\kappa}$. Kneip et al. (2015b) show that

$$\widehat{B}_{n,\kappa} = (2^\kappa - 1)^{-1} \left[\left(\widehat{\mu}_{n/2}^{(1)} + \widehat{\mu}_{n/2}^{(2)} \right) / 2 - \widehat{\mu}_n \right] \tag{3.33}$$

is a consistent estimator of the leading term of the bias, $Cn^{-\kappa}$, with an error of the same order as $R_{n,\kappa}$, i.e., $o(n^{-\kappa})$. Kneip et al. (2016) suggest randomly splitting the sample a large number of times, computing the bias estimate for each random split, and then averaging these estimates to obtain a bias estimate with less variance than an estimate based on only one sample-split.

Although the bias problem is solved by the generalized jackknife estimate of the bias, the remainder term $R_{n,\kappa} = o(n^{-\kappa})$ still creates problems when $\kappa < 1/2$. Kneip et al. (2015b) address this problem by randomly drawing a sub-sample $\mathcal{X}^*_{n_\kappa} \subset \mathcal{X}_n$ of size $n_\kappa = \lfloor n^{2\kappa} \rfloor < n$ when $\kappa < 1/2$ and computing [4]

$$\widehat{\mu}_{n_\kappa} = n_\kappa^{-1} \sum_{\{j \mid (X_j, Y_j) \in \mathcal{X}^*_{n_\kappa}\}} \widehat{\theta}(X_j, Y_j \mid \mathcal{X}_n). \tag{3.34}$$

The notation makes clear that each individual $\widehat{\theta}(X_j, Y_j \mid \mathcal{X}_n)$ is computed using the full sample $\mathcal{X}_n$, while the summation is over the observations in the subset of efficiency estimates of size n_κ.

Combining ideas, Kneip et al. (2015b) prove the following CLT.

Theorem 3.3. *Under appropriate regularity conditions given by Kneip et al. (2015b, 2026), for*

(i) $p+q \leq 3$ if the FDH estimator (and the CFDH estimator for estimating $\mathcal{C}(\Psi)$) is used,

(ii) $p+q \leq 4$ if Ψ exhibits CRS and the CFDH estimator is used,

(iii) $p+q \leq 4$ and if Ψ is convex and the DEA estimator (and the CDEA estimator for estimating $\mathcal{C}(\Psi)$) is used, or

(iv) $p+q \leq 5$ and if Ψ is convex and exhibits CRS and the DEA or the CDEA estimator is used,

$$\sqrt{n}\left(\widehat{\mu}_n - \mu_\theta - \widehat{B}_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2_\varphi) \tag{3.35}$$

as $n \to \infty$. When $\kappa < 1/2$,

$$\sqrt{n_\kappa}\left(\widehat{\mu}_{n_\kappa} - \mu_\theta - \widehat{B}_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma^2_\theta) \tag{3.36}$$

as $n \to \infty$.

As discussed in Kneip et al. (2015b, 2016, 2021, 2026), either (3.35) or (3.36) can be used in some cases. For example, when FDH estimators are used and $(p+q) = 3$, or when DEA estimators are used, CRS does not hold and $(p+q) = 4$, κ is close to $1/2$, but $\kappa < 1/2$. Kneip et al. (2015b) note that (3.36) is likely to provide a better approximation than (3.35) due to the order of the remainder term $R_{n,\kappa}$ that is omitted in both cases. Using the fact that part (iii) of Theorem 3.2 provides a consistent estimator of the variance σ^2_θ, the results in Theorem 3.3 can be used to build asymptotic confidence intervals for μ_θ by using the

[4] For $a \in \mathbb{R}$, $\lfloor a \rfloor$ denotes the largest integer that is less than or equal to a.

quantiles of the standard normal distribution. For example, if $\kappa < 1/2$, (3.36) can be used to obtain the asymptotically correct $1 - \alpha$ confidence interval for μ_θ given by

$$\left[\widehat{\mu}_{n_\kappa} - \widehat{B}_{n,\kappa} \pm z_{1-\alpha/2}\widehat{\sigma}_\theta/\sqrt{n_\kappa}\right], \tag{3.37}$$

where $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ and $\Phi^{-1}(\cdot)$ is the standard normal quantile function.

Remark 3.3. Data sharpening
It is well-known that estimation of the distribution of efficiency using the empirical distribution of DEA or FDH efficiency estimates is adversely affected in finite samples by the spurious mass of estimates equal to one (or zero in when directional efficiency estimates are used). However, in models appearing in the literature that develops statistical properties of non-parametric efficiency estimators, the DGP typically contains no mass at one. Smoothing the efficiency estimates near one may therefore help to improve results obtained with the CLTs described above. The idea is to sharpen the estimates $\widehat{\theta}(X_i, Y_i)$ *that fall in a small neighborhood of one by perturbing the estimates by some "small," continuous noise. The idea is used by Simar and Zelenyuk (2006) and Kneip et al. (2011) in bootstrap contexts. The adaptation to CLTs is proposed in Nguyen et al. (2022) and Simar et al. (2023), where Monte-Carlo results indicate substantial improvements of coverages for confidence intervals in small and moderate sample-size situations. By construction, the effects of sharpening disappear* $n \to \infty$.

Remark 3.4. Aggregate efficiency
The means analyzed so far are unweighted. In some applications, one might prefer to analyze weighted means where, for example, the weights may take into account the size or the economic importance of each producer. Simar and Zelenyuk (2018) and Simar et al. (2024) provide versions of the CLTs for these "aggregate efficiencies."

4 Robust Frontiers

By construction, FDH and DEA estimators involve estimates of Ψ that envelop all the data points in $\mathcal{X}_n$. Consequently, FDH and DEA estimators are sensitive to outliers that might incorporate errors in the data. Several methods have been proposed for detecting such points (e.g., see Wilson, 1993, 1995 and Simar, 2003), but another approach is to rely on a benchmark that is less sensitive to outliers than estimates of the support of (X, Y). To date, two such benchmarks, each with its own economic interpretation, exist in the literature, namely the order-m partial frontier (Cazals et al., 2002) and the order-*alpha* quantile frontier

(Aragon et al., 2005, and for the multivariate case, Daouia and Simar, 2007). The basic ideas are presented here while using the input orientation, but as with the discussion of FDH and DEA estimators, the ideas extend trivially to the output orientation. In addition, extensions to hyperbolic measures are developed by Wheelock and Wilson (2008) and Wilson (2011), and to directional distances by Simar and Vanhems (2012).

Recall that in (2.13) Ψ is described in terms of the support of $H_{XY}(x, y)$, the probability for a unit (x, y) to be dominated. Hence the input-oriented efficiency measure can be written as

$$\theta(x, y) = \inf\{\theta \mid H_{XY}(\theta x, y) > 0\}. \tag{4.1}$$

The main idea from Cazals et al. (2002) is that (for the input-oriented case) H_{XY} can be decomposed as

$$H_{XY}(x, y) = F_{X|Y}(x|y) S_Y(y), \tag{4.2}$$

where $F_{X|Y}(x \mid y) = \Pr(X \leq x \mid Y \geq y)$ and $S_Y(y) = \Pr(Y \geq y)$ (note the unusual inequality condition for Y in $F_{X|Y}(x \mid y)$). Then for all y such that $S_Y(y) > 0$,

$$\theta(x, y) = \inf\{\theta \mid F_{X|Y}(\theta x \mid y) > 0\}. \tag{4.3}$$

Therefore, the FDH estimator defined in (3.7) can be written as

$$\widehat{\theta}_{FDH}(x, y) = \inf\{\theta \mid \widehat{F}_{X|Y,n}(\theta x \mid y) > 0\} = \min_{i|Y_i \geq y} \left[\max_{j=1,\,\ldots,\,p} \left(\frac{X_{ij}}{x_j} \right) \right], \tag{4.4}$$

where $\widehat{F}_{X|Y,n}(\theta x \mid y)$ is the empirical analog of $F_{X|Y}(\theta x \mid y)$, i.e.,

$$\widehat{F}_{X|Y,n}(\theta x \mid y) = \frac{\sum_{i=1}^{n} \mathbb{1}(X_i \leq x,\ Y_i \geq y)}{\sum_{i=1}^{n} \mathbb{1}(Y_i \geq y)}. \tag{4.5}$$

It is easy to verify that (4.4) is equivalent to (3.8). This probabilistic formulation of the efficiency scores and their natural (FDH) non-parametric estimators, due to Cazals et al. (2002), is useful for describing robust frontiers that can serve as benchmarks for efficiency estimation, and also for describing conditional frontiers in Section 5.

4.1 Order-m Partial Frontiers

The economic idea behind the order-m frontier can be illustrated by first focusing on the case of one input for the input orientation. In this case, the full frontier (i.e., Ψ^{∂}, the frontier of Ψ) can be represented by the function

$$\varphi(y) = \inf\{x \mid F_{X|Y}(x|y) > 0\}. \tag{4.6}$$

Then, for any integer $m \geq 1$, the order-m cost frontier is defined by

$$\begin{aligned}\varphi_m(y) &= \mathbb{E}\left[\min(X_1,\ \ldots,\ X_m) Y \geq y\right] \\ &= \int_0^\infty \left[1 - F_{X|Y}(x,y)\right]^m dx = \varphi(y) + \int_{\varphi(y)}^\infty \left[1 - F_{X|Y}(x,y)\right]^m dx.\end{aligned} \tag{4.7}$$

From the first line of (4.7) it is apparent that $\varphi_m(y)$ provides a less extreme benchmark than the full frontier, because the unit operating at level (x, y) is benchmarked against the expectation of the minimum input among m randomly drawn firms producing more output than y. The equivalences in (4.7) are rather straightforward and show clearly that if $m \to \infty$, $\varphi_m(y) \to \varphi(y)$, where $\varphi(y)$ is the full frontier. Note also that an order-m efficiency score is given here by $\theta_m(x, y) = \varphi_m(y)/x$, which is not bounded above by one unless $m \to \infty$.

Cazals et al. (2002) extend the concept to the multivariate input case. Consider again m random draws X_i, $i = 1, \ldots, m$ from the conditional distribution $F_{X|Y}(\cdot \mid y)$, and define the random set

$$\Psi_m(y) = \bigcup_{i=1}^{m} \{(u, v) \mid u \geq X_i,\ v \geq y\}. \tag{4.8}$$

Then for any x, define

$$\widetilde{\theta}_m(x, y) = \inf\{\theta \mid (\theta x, y) \in \Psi_m(y) = \min_{i=1,\ \ldots,\ m} \left\{ \max_{j=1,\ \ldots,\ p} \frac{X_{ij}}{x_j} \right\}. \tag{4.9}$$

Since the set $\Psi_m(y)$ being random, $\widetilde{\theta}_m(x, y)$ is random. Cazals et al. (2002) define the order-m efficiency score, in the general multidimensional case as

$$\theta_m(x, y) = \mathbb{E}(\widetilde{\theta}_m(x, y) \mid Y \geq y), \tag{4.10}$$

i.e., as the expectation of $\widetilde{\theta}_m(x, y)$. They prove also that

$$\theta_m(x, y) = \int_0^\infty \left[1 - F_{X|Y}(ux \mid y)\right]^m du = \theta(x, y) + \int_{\theta(x,y)}^\infty \left[1 - F_{X|Y}(ux \mid y)\right]^m du. \tag{4.11}$$

The integrals in (4.11) are univariate in u and again, $\theta_m(x, y) \to \theta(x, y)$ as $m \to \infty$. These expressions suggest a natural, non-parametric estimator of $\theta_m(x, y)$ obtained by replacing the integrals with the empirical version of $F_{X|Y}$ to obtain

$$\widehat{\theta}_m(x, y) = \int_0^\infty \left[1 - \widehat{F}_{X|Y,n}(ux \mid y)\right]^m du = \widehat{\theta}(x, y) + \int_{\widehat{\theta}(x,y)}^\infty \left[1 - \widehat{F}_{X|Y,n}(ux \mid y)\right]^m du, \tag{4.12}$$

where $\widehat{\theta}(x, y)$ is the FDH estimator of the efficiency score. Cazals et al. (2002) propose using the Monte-Carlo method to compute these integrals. However, the methods of Daraio et al. (2020) for estimation of directional distances extend trivially to the input or output orientation, providing an analytical approach to determine the exact value of the integrals in (4.11) by using the fact that $\widehat{F}_{X|Y,n}(\cdot \mid y)$ is discrete and jumps only at the observations X_i where $Y_i \geq y$. Moreover, the method of Daraio et al. (2020) is faster than the Monte Carlo method suggested by Cazals et al. (2002). Order-m efficiency estimators are extended to hyperbolic directions by Wilson (2011), and to the additive, directional case by Simar and Vanhems (2012).

Asymptotic Properties

Cazals et al. (2002) derive asymptotic properties of the estimator in (4.12). For fixed, finite m, $\widehat{\theta}_m(x,y)$ achieves a parametric $\sqrt{n}$ rate of convergence with a limiting normal distribution, i.e.,

$$\sqrt{n}\left(\widehat{\theta}_m(x,y) - \theta_m(x,y)\right) \xrightarrow{\mathcal{L}} N\left(0, \sigma^2_m(x,y)\right), \tag{4.13}$$

where an expression for $\sigma^2_m(x,y)$ is given by Cazals et al.. This result enables inference on $\theta_m(x,y)$ using standard bootstrap methods. In addition, if $m = m(n) \to \infty$ at rate $n \log n S_Y(y)$ as $n \to \infty$, then

$$n^{1/(p+q)}\left(\widehat{\theta}_{m(n)}(x,y) - \theta(x,y)\right) \xrightarrow{\mathcal{L}} \text{Weibull}(\eta(x,y)). \tag{4.14}$$

In this case, the order-m estimator behaves exactly as does the ordinary FDH estimator; e.g., see (3.25). As m increases the $\sqrt{n}$ rate of convergence is lost, and the curse of the dimensionality re-appears. However, in finite samples, the order-m estimator is more robust than the FDH estimator provided m is not too large, since the corresponding frontier estimate does not envelop all the data points.

The last suggests a strategy for choosing m in practice. If we do not want to fix m at a some predetermined values fixed by the practitioner, the usual strategy is to compute the estimator for several values of m (e.g., 25, 50, 100, ...) and compute for each value of m the percentage of data points lying outside the order-m frontier. These points are identified by values $\widehat{\theta}_m(X_i, Y_i) > 1$. This percentage decreases to zero when m increases. So we might choose m by fixing a desired fixed percentage of points outside the frontier (e.g., 5 percent), or by looking where the plot of these percentages versus the corresponding values of m present an *elbow effect*. From that value of m, the percentage of points outside the order-m frontier begins to decrease less quickly, indicating that the remaining points outside the frontier are extreme and potentially outliers (see Simar, 2003 and Daouia and Gijbels, 2011). Daouia et al. (2012) show how the order-m frontier can be viewed as a regularized estimator of the full frontier, if m increases to ∞ but at the slower rate, reflecting the bias in the estimator of the full frontier. This bias can be consistently estimated under mild regularity conditions.

4.2 Order-α Quantile Frontiers

Aragon et al. (2005) introduce the order-α quantile frontier for a univariate response variable (e.g., for a univariate input for the input orientation). The extension to the fully multivariate case is developed by Daouia and Simar (2007). The idea is rather simple, and involves replacing the benchmark given by the full minimal support of X given $Y \geq y$, as in (4.3), with the benchmark given by the input level not exceeded by $(1-\alpha) \times 100$-percent of firms in the population of units producing at least the output level y. This leads to defining the order-α (input) efficiency measure as

$$\theta_\alpha(x, y) = \inf \left\{ \theta \mid F_{X|Y}(\theta x \mid y) > 1 - \alpha \right\}. \tag{4.15}$$

If $\theta_\alpha(x, y) = 1$, the unit (x, y) is input efficient at level $\alpha \times 100\%$ because it is dominated by other units (less input than x and more output than y) with a probability $1 - \alpha$. Clearly, if $\alpha \to 1$, then $\theta_\alpha(x, y) \to \theta(x, y)$.

A natural non-parametric estimator of $\theta_\alpha(x, y)$ is provided by plugging the empirical version of $F_{X|Y}$ into (4.15), yielding

$$\widehat{\theta}_\alpha(x, y) = \inf \left\{ \theta \mid \widehat{F}_{X|Y,n}(\theta x \mid y) > 1 - \alpha \right\}. \tag{4.16}$$

Daouia and Simar (2007) provide a simple algorithm for practical computation. The output orientation precedes along similar lines. Wheelock and Wilson (2008)) extend the concept to hyperbolic measures, and Simar and Vanhems (2012) extend the idea to directional distances. See also Daraio et al. (2020).

Asymptotic Properties

Asymptotic properties of $\theta_\alpha(x, y)$ are similar to the those of the order-m efficiency estimator. For example, in the input orientation

$$\sqrt{n} \left(\widehat{\theta}_\alpha(x, y) - \theta_\alpha(x, y) \right) \xrightarrow{\mathcal{L}} N(0, \sigma_\alpha^2(x, y)), \tag{4.17}$$

with an explicit expression for $\sigma_\alpha^2(x, y)$ given in Daouia and Simar (2007). This result allows allows one to make inference on $\theta_\alpha(x, y)$ by standard bootstrap methods. In addition, if

$\alpha = \alpha(n) \to 1$ fast enough as $n \to \infty$, then

$$n^{1/(p+q)} \left(\widehat{\theta}_{\alpha(n)}(x,y) - \theta(x,y) \right) \xrightarrow{\mathcal{L}} \text{Weibull}(\eta(x,y)) \tag{4.18}$$

(see details in Daouia and Simar, 2007). As with the order-m estimator, in this case the root-n convergence is lost and the curse-of-dimensionality arises.

5 Heterogeneity and Environmental Factors

The economic and statistical models discussed so far focus on inputs $X \in \mathbb{R}_+^p$ and outputs $Y \in \mathbb{R}_+^q$. However, in some situations, external factors such as ownership type, weather conditions, geographical considerations, etc. may significantly impact the production process. Failing to incorporate these variables into the model can lead to misleading and invalid results on producers' performances.

Let $Z \in \mathbb{R}^d$ denote a random vector of "environmental" variables that may affect how units convert inputs X into outputs Y. Let

$$f_{XYZ}(x,y,z) = f_{XY|Z}(x,y \mid z) f_Z(z). \tag{5.1}$$

be the joint density of (X,Y,Z) with support $\mathcal{P} \subset \mathbb{R}_+^p \times \mathbb{R}_+^q \times \mathbb{R}^d$. Denote the support of $f_Z(z)$ by $\mathcal{Z}$, and denote the support of $f_{XY|Z}$ by Ψ^z. Then Ψ^z is the set of feasible combinations of inputs and outputs for a firm facing the environmental conditions $Z = z$, i.e.,

$$\Psi^z = \{(x,y) \mid x \text{ can produce } y \text{ when } Z = z\}. \tag{5.2}$$

Necessarily, Z can affect the production process either (i) only through Ψ^z, the support of (X,Y) when $Z = z$, or (ii) only through the density $f_{XY|Z}(x,y \mid z)$, thereby affecting the probability for a firm to be near its optimal boundary, or (iii) through both Ψ^z and $f_{XY|Z}(x,y \mid z)$. By construction,

$$\Psi = \bigcup_{z \in \mathcal{Z}} \Psi^z. \tag{5.3}$$

Therefore, in general, $\Psi^z \subseteq \Psi \ \forall \ z \in \mathcal{Z}$ and $\Psi \subset \mathbb{R}_+^{p+q}$. Whether Ψ is a feature of the underlying DGP, or whether Ψ is useful for benchmarking the performance of units, depends on the "separability" condition described by Simar and Wilson (2007) that is discussed below in Section 5.2.

5.1 General Cases: Conditional Frontier Models

In general, the relevant benchmark when facing environmental factors $Z = z$ is given by the efficient boundary of Ψ^z, i.e.,

$$\Psi^{z\partial} = \left\{ (x, y) \mid (x, y) \in \Psi^z,\ (\gamma x, \gamma^{-1} y) \notin \Psi^z \text{ for any } \gamma \in (0, 1) \right\}. \tag{5.4}$$

The corresponding conditional measure of efficiency (in the input orientation) is

$$\theta(x, y \mid z) = \inf \left\{ \theta \mid (\theta x, y) \in \Psi^z \right\}. \tag{5.5}$$

Note that Ψ^z can be viewed as the support of $H_{XY|Z}(x, y \mid z) = \Pr(X \leq x,\ Y \geq y \mid Z = z)$. Consequently,

$$\theta(x, y \mid z) = \inf \left\{ \theta \mid H_{XY|Z}(\theta x, y \mid z) > 0 \right\}. \tag{5.6}$$

In the input orientation, for any y where $S_{Y|Z}(y \mid z) = \Pr(Y \geq y \mid Z = z) > 0$, (5.6) can be written as

$$\theta(x, y \mid z) = \inf \left\{ \theta \mid F_{X|YZ}(\theta x \mid y, z) > 0 \right\}, \tag{5.7}$$

where $F_{X|YZ}(\theta x \mid y, z) = \Pr(X \leq x \mid Y \geq y, Z = z)$, noting the inequality for the condition on Y but the equality for Z.

Given a sample $\mathcal{S}_n = \{(X_i, Y_i, Z_i)\}_{i=1}^n$, a natural, non-parametric estimator of $F_{X|YZ}$ is the kernel estimator

$$\widehat{F}_{X|YZ,n}(x \mid y, z) = \frac{\sum_{i=1}^n \mathbb{1}(X_i \leq x,\ Y_i \geq y) K\left(\frac{Z_i - z}{h}\right)}{\sum_{i=1}^n \mathbb{1}(Y_i \geq y) K\left(\frac{Z_i - z}{h}\right)},, \tag{5.8}$$

where $h = (h_1, \ldots, h_d)$ is a vector of d bandwidths and $K(\cdot)$ d-variate kernel function. As noted by Daraio and Simar (2007b), kernels with bounded support (e.g., the Epanechnikov, Quartic, Triweight, or other kernels) must be used. Bandwidths can be optimized by least-squares cross-validation (LSCV), which provides bandwidths with the optimal order $O\left(n^{-1/(d+4)}\right)$. See Silverman (1986) or Li et al. (2013) for details and discussion.

Regardless of the choice of kernel function, the conditional FDH estimator is obtained by replacing $F_{X|YZ}(\theta x \mid y, z)$ in (5.6) with the estimator in (5.8) to obtain

$$\widehat{\theta}_{FDH}(x, y \mid z) = \inf \left\{ \theta \mid \widehat{F}_{X|YZ,n}(\theta x \mid y, z) > 0 \right\} = \min_{\{i \mid Y_i \geq y,\ ||Z_i - z|| \leq h\}} \ \max_{j=1,\ \ldots,\ p} \left(\frac{X_i^j}{x^j} \right), \tag{5.9}$$

where $||a||$ denotes the Euclidean norm of a vector a. For continuous Z, the estimator in (5.9) is a localized version of the ordinary FDH estimator, localized for data such that Z_i is in a neighborhood of z.

Daraio and Simar (2007b) provide a conditional DEA estimator, similar to the conditional FDH estimator in (5.9), but based on a convex version of the local set of reference points. The conditional DEA estimator can be computed as

$$\widehat{\theta}_{DEA}(x, y \mid z) = \inf \left\{ \theta \mid y \leq \sum_{i \in \mathcal{I}_z} \xi_i Y_i,\ \theta x \geq \sum_{i \in \mathcal{I}_z} \xi_i X_i,\ \sum_{i \in \mathcal{I}_z} \xi_i = 1,\ \xi_i \geq 0\ \forall\, i \in \mathcal{I}_z \right\}, \quad (5.10)$$

where $\mathcal{I}_z = \{i \mid ||Z_i - z|| \leq h\}$.

Remark 5.1. The case where Z is discrete
When Z is discrete or qualitative, the situation becomes simpler. The localization described earlier reduces to treating each possible value of Z independently, with each value defining the set of local observations (X_i, Y_i). In other words, Z serves to divide the data into groups corresponding to each observed value of Z, and with each group treated independently. However, the number of observations in each local problem may be too small to obtain meaningful estimates. In such cases, some groups should be merged.

The asymptotic properties for the individual estimators of conditional efficiencies are derived by Jeong et al. (2010). To summarize, these properties are analogous to those for the unconditional case with n replaced by $n_h = nh^d$, which is the order of the number of data points in the neighborhood of z. When using LSCV- optimal bandwidths, the number of such data points reduces to an order of $n^{4/(4+d)}$. This reduces the rate of the unconditional envelopment estimator when d increases. For instance, for an estimator that has a rate n^{κ} in the unconditional case, the rate decreases to $n^{4\kappa/(4+d)}$ for the conditional case. Cazals et al. (2025) propose a tentative solution to this disappointing fact by employing single index models for Z that reduces the effect of Z to one dimension.

Central limit theorems for means of conditional efficiency measures, analogous to Theorems 3.2-3.3 presented above, have been derived by Daraio et al. (2018). Their presentation requires some technical details, but the procedure for correcting the bias of the conditional estimators is based on a jackknife method similar to the method described above for the unconditional case (see details in Daraio et al. 2018, Theorems 4.1–4.4).

Remark 5.2. Robust conditional frontiers
The order-m and order-α quantile robust frontiers can be adapted to the conditional setup. For radial distances, see Cazals et al. (2002) and Daouia and Simar (2007), and for directional distances, see Simar and Vanhems (2012). Daraio et al. (2020) provide methods for practical computation in the later case. In summary, these robust versions retain the same

properties as their unconditional counterparts, such as robustness to extreme values and outliers and improved asymptotic properties. Moreover, they achieve asymptotic normality at a rate not affected by the dimension of $p+q$, but only by the dimension of Z, i.e., at rate $n^{4/(4+d)}$.

5.2 Separability Condition and Two-Stage Models

A large number of papers in the literature use two-stage procedures to analyze the effect of environmental variables Z on the production process. These analyses involve first estimating unconditional efficiency using FDH or DEA estimators, and then regressing the efficiency estimates on Z in a limited dependent variable model or sometimes in a simple linear regression.

These approaches are problematic for several reasons. First, it is clear that the first-stage efficiency estimates provide meaningful measures if and only if the support of Ψ^z is not affected by z. Simar and Wilson (2007) call this the "separability condition," and note that this is a restrictive assumption that should be tested. Formally, this condition assumes that

$$\Psi^z = \Psi \ \forall \ z \in \mathcal{Z}, \tag{5.11}$$

allowing the joint support of (X, Y, Z) to be factorized as

$$\mathcal{P} = \Psi \times \mathcal{Z}. \tag{5.12}$$

The assumption implies that, in the input orientation for example, $\theta(x, y \mid z) = \theta(x, y)$ for all $z \in \mathcal{Z}$. As discussed by Simar and Wilson (2007, 2011b), if this is not true, then $\theta(x, y)$ and the unconditional, first-stage efficiency estimates have no particular economic nor statistical meaning since the unit (x, y) is benchmarked against a frontier that is not in general attainable given the environmental condition described by z. If the separability does not hold, it is better in the first stage to use the conditional measures of efficiency and analyze how these are affected by Z in some appropriate model (e.g., see Bădin et al. (2012, 2014 and Daraio and Simar, 2014).

Second, even if the separability condition (5.11) holds, the second stage estimation is complicated by the fact that one does not have the true efficiencies on the left-hand side of whatever regression model is used, but instead only *estimates* of the true efficiencies. Moreover, these estimates are biased and typically have slow convergence rates. Consequently, standard methods of interest (which are invariably used in two-stage models in the efficiency literature) cannot provide valid inference as discussed by Simar and Wilson (2007) and in

Kneip et al. (2015b, Section 5). For cases where the separability condition holds, Fève et al. (2023) suggest an alternative to the second stage regression that avoids the failure of the standard regression methods. But again, the separability assumption in (5.11) is a strong assumption, and it is often rejected when tested. Daraio et al. (2018) provide a test of the separability condition as discussed below in Section 6.4.

6 Testing Issues

Kneip et al. (2016) show have shown how the various CLTs developed by Kneip et al. (2015b) enable the construction of test statistics, with simple, known asymptotic distributions under the relevant null hypothesis, that can be used to test various hypotheses on the structure of the underlying DGP. Simulation results provided by Kneip et al. (2016) indicate that the tests have good size and power properties in finite-sample situations.

The simplest case involves testing equivalent versus different expected efficiencies across two independent groups of firms. The other tests described below require splitting the original sample into two independent sub-samples in order to derive the asymptotic distribution of the test statistic under the null hypothesis. Below, the simplest case is presented first. However, as explained below, the empirical researcher must first verify assumptions about the shape of the attainable sets (i.e., convex or non-convex, CRS or non-CRS) to ensure that appropriate estimators are used.

6.1 Testing the Equality of Means of Efficiencies

Suppose there are two groups G_1 and G_2 of firms defined by some criterion. In addition, suppose the researcher has two independent samples, one from each group, of sizes n_1 and n_2, respectively. A natural research question is whether the two groups have the same expected or mean efficiencies, i.e., whether $\mu_{1,\theta} = \mu_{2,\theta}$ where $\mu_{\ell,\theta} = \mathbb{E}(\theta(X,Y) \mid (X,Y) \in G_\ell)$ for $\ell \in \{1,\ 2\}$. In other words, it is desired to test the null hypothesis

$$H_0\colon \mu_{1,\theta} = \mu_{2,\theta} \tag{6.1}$$

versus the alternative, one-sided hypothesis

$$H_1\colon \mu_{1,\theta} > \mu_{2,\theta}. \tag{6.2}$$

Note that the labels on the two groups are arbitrary and can be reversed. In addition, use of a two-sided alternative $H_1\colon \mu_{1,\theta} \neq \mu_{2,\theta}$ follows reasoning similar to that below. Note also that here, no assumption on whether the two groups face the same frontier is imposed. Kneip et al. (2016, Section 3.1.2) show how the test discussed here can be adapted to incorporate the assumption of a common frontier for the two groups.

Let $\mathcal{X}_{\ell,n_\ell} = \{(X_i, Y_i)\}_{i=1}^{n_\ell}$ denote the iid samples of firms in groups $\ell \in \{1,\ 2\}$, and let

$$\widehat{\mu}_{\ell,\theta} = n_\ell^{-1} \sum_{(X_i,Y_i)\in\mathcal{X}_{\ell,n_\ell}} \widehat{\theta}(X_i, Y_i|\mathcal{X}_{\ell,n_\ell}) \tag{6.3}$$

be the sample means of estimated efficiencies for each group. Here, depending on the underlying economic assumptions on the attainable sets Ψ_ℓ, the estimators $\widehat{\theta}(\cdot)$ can be either the FDH, DEA, CFDH or CDEA estimators with corresponding convergence rate n^κ with values of κ given in Table 1.

Theorem 3.3 applies for each of the two groups. For each group $\ell \in \{1,\ 2\}$ the bias of the corresponding sample mean in (6.3) is estimated by $\widehat{B}_{\ell,\kappa,n_\ell}$ as described in Section 3.3.3. A test statistics can then be derived easily due to the independence between the two samples. This independence plays a crucial role, and avoids complications due to covariances.

The results of Theorem 3.3 make clear that one must carefully account for the value of $(p+q)$. In situations where (3.35) applies, it follows that

$$T_{1,n_1,n_2} = \frac{(\widehat{\mu}_{1,n_1} - \widehat{\mu}_{2,n_2}) - \left(\widehat{B}_{1,\kappa,n_1} - \widehat{B}_{2,\kappa,n_2}\right) - (\mu_{1,\theta} - \mu_{2,\theta})}{\sqrt{\frac{\widehat{\sigma}^2_{1,\theta,n_1}}{n_1} + \frac{\widehat{\sigma}^2_{2,\theta,n_2}}{n_2}}} \xrightarrow{\mathcal{L}} N(0,1), \tag{6.4}$$

provided $n_1/n_2 \to c > 0$ as $n_1, n_2 \to \infty$, where c is a constant and where the $\widehat{\sigma}^2_{\ell,\theta,n_\ell} = n_\ell^{-1}\sum_{i=1}^{n_\ell}\left[\widehat{\theta}(X_{\ell,i}, Y_{\ell,i} \mid \mathcal{X}_{\ell,n_\ell}) - \widehat{\mu}_{\ell,n_\ell}\right]^2$ are consistent estimators of $\sigma^2_{\ell,\theta} = \mathbb{V}(\theta(X,Y) \mid (X,Y) \in G_\ell)$ for $\ell \in \{1,\ 2\}$. In the more frequent case where $p+q$ is large and $\kappa < 1/2$, (3.36) applies, requiring the means

$$\widehat{\mu}_{\ell,n_{\ell,\kappa}} = n_{\ell,\kappa}^{-1} \sum_{(X_{\ell,i},Y_{\ell,i})\in\mathcal{X}^*_{\ell,n_{\ell,\kappa}}} \widehat{\theta}\left(X_{\ell,i}, Y_{\ell,i}) \mid \mathcal{X}_{\ell,n_\ell}\right), \tag{6.5}$$

to be computed over random sub-samples $\mathcal{X}^*_{\ell,n_{\ell,\kappa}}$ of size $n_{\ell,\kappa} = \lfloor n_\ell^{2\kappa} \rfloor$ for each of the two groups $\ell\{1,\ 2\}$. The notation in (6.5) makes clear that each individual efficiency estimator is computed with the full samples $\mathcal{X}_{\ell,n_\ell}$, but the summation is only over the elements of $\mathcal{X}^*_{\ell,n_{\ell,\kappa}}$. Then it follows that

$$T_{2,n_{1,\kappa},n_{2,\kappa}} = \frac{\left(\widehat{\mu}_{1,n_{1,\kappa}} - \widehat{\mu}_{2,n_{2,\kappa}}\right) - \left(\widehat{B}_{1,\kappa,n_1} - \widehat{B}_{2,\kappa,n_2}\right) - (\mu_{1,\theta} - \mu_{2,\theta})}{\sqrt{\frac{\widehat{\sigma}^2_{1,\theta,n_1}}{n_{1,\kappa}} + \frac{\widehat{\sigma}^2_{2,\theta,n_2}}{n_{2,\kappa}}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{6.6}$$

provided $n_1/n_2 \to c > 0$ as $n_1, n_2 \to \infty$,

Under the null hypothesis, $\mu_{1,\theta} - \mu_{2,\theta} = 0$. Therefore, depending on the chosen estimator and on the value of $p+q$, one of the two statistics in (6.4) or (6.6) can be computed and compared with the appropriate quantile of the standard normal distribution to obtain an asymptotically correct p-value for H_0. Alternatively, (6.4) or (6.6) can be used with quantiles of the standard normal distribution to derive confidence intervals for $\mu_{1,\theta} - \mu_{2,\theta}$ with asymptotically correct coverage.

6.2 Testing Convexity versus non-Convexity of Ψ

As discussed earlier, FDH estimators do not require convexity of Ψ and are consistent under only the free disposability assumption and other mild regularity conditions, but DEA estimators require both convexity and free disposability. In the empirical literature, DEA estimators are used much more frequently than FDH estimators, with apparently little regard to whether convexity of Ψ holds. Of course, one should test the null hypothesis $H_0\colon \Psi$ is convex versus the alternative hypothesis $H_1\colon \Psi$ is not convex. Kneip et al. (2016) provide a test that is easy to implement.

Working in the input orientation (for illustrative purposes), by construction $\widehat{\theta}_{\text{DEA}}(X_i, Y_i \mid \mathcal{X}_n) \le \widehat{\theta}_{\text{FDH}}(X_i, Y_i \mid \mathcal{X}_n)$. Therefore, $\widehat{\mu}_{\text{DEA},n} \le \widehat{\mu}_{\text{FDH},n}$, where $\widehat{\mu}_{\bullet,n} = n^{-1}\sum_{i=1}^{n} \widehat{\theta}_{\bullet}(X_i, Y_i \mid \mathcal{X}_n)$ with $\bullet$ denoting either the FDH or the DEA case. Under the null hypothesis of convexity, both means estimate consistently the same quantity, but if Ψ is not convex we expect large difference in the two sample means. Consequently, H_0 should be rejected if the difference is "large". But, some complication arises as explained by Kneip et al. (2016). Due to the covariances between $\widehat{\theta}_{\text{DEA}}(X_i, Y_i \mid \mathcal{X}_n)$ and $\widehat{\theta}_{\text{FDH}}(X_i, Y_i \mid \mathcal{X}_n)$, $n^{\gamma}(\widehat{\mu}_{\text{FDH},n} - \widehat{\mu}_{\text{DEA},n})$ converges under the null to a degenerate distribution for any $\gamma \le 1/2$. Therefore the two sample means must be computed over two independent samples.

One way to accomplish this is to first randomly split $\mathcal{X}_n$ into two independent sub-samples $\mathcal{X}_{n_1}$ and $\mathcal{X}_{n_2}$ with $n_1 + n_2 = n$ and then use DEA estimators for observations in $\mathcal{X}_{n_1}$, with rate of convergence n^{κ}, $\kappa_1 = 2/(p+q+1)$, and use FDH estimators for observations in $\mathcal{X}_{n_2}$, with rate involving $\kappa_2 = 1/(p+q)$. Kneip et al. (2016) suggest balancing the values of n_1 and n_2 so that the rates are similar in both sub-samples, i.e., so that $n_1^{\kappa_1} \approx n_2^{\kappa_2}$ with $n_1 + n_2 = n$. Then the means to be compared are computed separately and independently using the sub-samples $\mathcal{X}_{n_1}$ and $\mathcal{X}_{n_2}$, yielding $\widehat{\mu}_{\text{DEA},n_1}$ and $\widehat{\mu}_{\text{FDH},n_2}$ (respectively). Due to the independence between the two sub-samples, the remainder of the procedure is analogous to the one described above for testing the equality of mean efficiencies across two groups of producers.

Kneip et al. (2016) show that under the null hypothesis of convexity, provided $\kappa_2 \geq 1/2$,

$$T_{3,n} = \frac{\widehat{\mu}_{\mathrm{FDH},n_2} - \widehat{\mu}_{\mathrm{DEA},n_1} - (\widehat{B}_{\mathrm{FDH},n_2,\kappa_2} - \widehat{B}_{\mathrm{DEA},n_1,\kappa_1})}{\sqrt{\frac{\widehat{\sigma}^2_{\mathrm{DEA},n_1}}{n_1} + \frac{\widehat{\sigma}^2_{\mathrm{FDH},n_2}}{n_2}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{6.7}$$

as $n \to \infty$, where the bias and variance estimators for both cases are obtained using the procedure described in Section 3.3.3.

Alternatively, if $\kappa_2 > 1/2$ then

$$T_{4,n} = \frac{\widehat{\mu}_{\mathrm{FDH},n_{2,\kappa}} - \widehat{\mu}_{\mathrm{DEA},n_{1,\kappa}} - (\widehat{B}_{\mathrm{FDH},n_2,\kappa_2} - \widehat{B}_{\mathrm{DEA},n_1,\kappa_1})}{\sqrt{\frac{\widehat{\sigma}^2_{\mathrm{DEA},n_1}}{n_{1,\kappa}} + \frac{\widehat{\sigma}^2_{\mathrm{FDH},n_2}}{n_{2,\kappa}}}} \xrightarrow{\mathcal{L}} N(0,1), \tag{6.8}$$

as $n \to \infty$, where $\kappa = \kappa_2$, $n_{\ell,\kappa} = \lfloor n_\ell^{2\kappa} \rfloor$ for $\ell \in \{1,\ 2\}$ and, as in Theorem 3.3, the means are computed over the random sub-sample of sizes $n_{\ell,\kappa}$, $\ell \in \{1,\ 2\}$.

The null hypothesis of convexity of Ψ is rejected whenever $1 - \Phi(T_{3,n})$, or $1 - \Phi(T_{4,n})$ (depending on the value of $p+q$) is less than α (e.g., $\alpha = .1$, .05 or .01).

The procedure as described by Kneip et al. (2016) is based on a single random split of the original $\mathcal{X}_n$, and this may introduce some ambiguity and arbitrariness. Different splits could provide different decision. To remove this ambiguity, Simar and Wilson (2020, 2026) suggest methods for combining results from multiple sample-splits that lead to consistent, valid tests as described below in Section 6.5.

6.3 Testing CRS versus Non-CRS

Regardless of the outcome of the test of CRS versus non-CRS, one must decide whether to use the FDH or CFDH estimators, or the DEA or CDEA estimators to estimate efficiency. Thus it is useful to test CRS versus non-CRS, i.e., to test the null hypothesis $H_0\colon \Psi = \mathcal{C}(\Psi)$ against the alternative hypothesis $H_1\colon \Psi \subset \mathcal{C}(\Psi)$. If convexity holds, then under the null hypothesis of CRS, both the DEA and the CDEA are consistent estimators of $\theta(X,Y)$, but under the alternative, only the DEA estimator is consistent. Kneip et al. (2016) give an appropriate test for cases where Ψ is convex. If convexity does not hold, then under H_0, both the FDH and the CFDH are consistent estimators of $\theta(X,Y)$, but under the alternative, only the FDH estimator is consistent. For this case, a test of CRS versus non-CRS is provided by Kneip et al. (2026).

For purposes of illustration, assume that Ψ is convex. The test statistics provided by Kneip et al. (2016) can be used, and these are based on the fact that (for the input orientation) $\widehat{\theta}_{\mathrm{CDEA}}(X_i, Y_i \mid \mathcal{X}_n) \leq \widehat{\theta}_{\mathrm{DEA}}(X_i, Y_i \mid \mathcal{X}_n)$ and hence the sample means of the DEA and

CDEA estimates computed over the entire sample $\mathcal{X}_n$ are such that their arithmetic means computed over the full sample $\mathcal{X}_n$ are such that $\widehat{\mu}_{\text{DEA},n} - \widehat{\mu}_{\text{CDEA},n} \geq 0$. Under the null, this difference is expected to be "small," and under the alternative, it is expected to be "large". Unfortunately, for reasons similar to those involving covariances between the estimators in the test of convexity, $n^{\gamma}(\widehat{\mu}_{\text{FDH},n} - \widehat{\mu}_{\text{DEA},n})$ converges under the null to a degenerate distribution for any $\gamma \leq 1/2$. The solution here is again to first randomly split the full sample $\mathcal{X}_n$ into two independent sub-samples $\mathcal{X}_{n_1}$ and $\mathcal{X}_{n_2}$ of sizes n_1 and n_2, then use the DEA estimator to estimate efficiency among the observations in $\mathcal{X}_{n_1}$, and the CDEA estimator for the observations in $\mathcal{X}_{n_2}$, and then finally compute the corresponding means within each sub-sample. Kneip et al. (2016) prove that under the null hypothesis of CRS, both the DEA and CDEA estimators share the same rate with $\kappa = 2/(p+q)$, so the splitting can be balanced with $n_1 \approx n_2$, $n_1 + n_2 = n$.

Here, as in the test of equivalent mean efficiencies, Kneip et al. (2016) show that the appropriate test statistic depends on the value of $p+q$. If $p+q \leq 5$, then under the null hypothesis of CRS,

$$T_{5,n} = \frac{(\widehat{\mu}_{\text{DEA},n_1} - \widehat{\mu}_{\text{CDEA},n_2}) - \left(\widehat{B}_{\text{DEA},\kappa,n_1} - \widehat{B}_{\text{CDEA},\kappa,n_2}\right)}{\sqrt{\frac{\widehat{\sigma}^2_{\text{DEA},n_1}}{n_1} + \frac{\widehat{\sigma}^2_{\text{CDEA},n_2}}{n_2}}} \xrightarrow{\mathcal{L}} N(0,1), \tag{6.9}$$

as $n \to \infty$, where the bias and variance estimators for both the DEA and CDEA cases are obtained by the procedure described in Section 3.3.3. Alternatively, if $p+q > 5$, the sample means must be computed over random sub-samples of $\mathcal{X}_{n_\ell}$ of size $n_{\ell,\kappa} = \lfloor n_\ell^{2\kappa} \rfloor$, $\ell \in \{1,\ 2\}$. Then under the null,

$$T_{6,n} = \frac{(\widehat{\mu}_{\text{VRS},n_{1,\kappa}} - \widehat{\mu}_{\text{CRS},n_{2,\kappa}}) - \left(\widehat{B}_{\text{VRS},\kappa,n_1} - \widehat{B}_{\text{CRS},\kappa,n_2}\right)}{\sqrt{\frac{\widehat{\sigma}^2_{\text{VRS},n_1}}{n_{1,\kappa}} + \frac{\widehat{\sigma}^2_{\text{CRS},n_2}}{n_{2,\kappa}}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{6.10}$$

as $n \to \infty$, The null hypothesis of CRS whenever $1 - \Phi(T_{5,n})$, or $1 - \Phi(T_{6,n})$ (depending on the value of $p+q$) is less than the nominal test size α. The ambiguity resulting from a single random sample-split can be removed as described below.

A similar procedure for testing CRS for general technologies, without imposing convexity, is provided by Kneip et al. (2026). The procedure works along the same lines as the test described here while FDH and CFDH estimators instead of DEA and CDEA estimators. An important difference, however, is established by Kneip et al. (2026). Even under the null, the FDH and CFDH estimators do not share the same convergence rate, as shown in Table

1 where $\kappa_1 = 1/(p+q)$ when using the FDH estimator and $\mathcal{X}_{n_1}$, while $\kappa_2 = 1/(p+q-1)$ when using the CFDH estimator and $\mathcal{X}_{n_2}$. See Kneip et al. (2026) for details and discussion.

6.4 Testing the Separability Condition

Daraio et al. (2018) provide a method for testing the separability condition (5.11), i.e., to test the null hypothesis $H_0\colon \Psi^z = \Psi \ \forall \ z \in \mathcal{Z}$ against the alternative hypothesis $H_1\colon \Psi^z \subset \Psi$ for some $z \in \mathcal{Z}$. The main idea is to build a test statistic comparing the expectation of the unconditional efficiencies $\mu = \mathbb{E}(\theta(X,Y))$ with the expectation of the conditional efficiencies $\mu_c = \mathbb{E}(\theta(X,Y \mid Z)$. By construction (for the input orientation, chosen here for illustration), $\mu_c - \mu \geq 0$, with equality holding if and only if the separability condition in H_0 holds. So here again, as with the tests of convexity and CRS, a one-sided test can be built. The test rejects H_0 if the difference between two independent estimators of μ_c and μ is "large". The CLTs developed by Daraio et al. (2018) for conditional efficiency estimators allow construction of the appropriate test statistics, including bias corrections for each estimator.

The separability test also requires splitting the original sample $\mathcal{X}_n$ into two independent sub-samples $\mathcal{X}_{n_1}$ and $\mathcal{X}_{n_2}$ with $n_1 + n_2 = n$ as in the tests described in Sections 6.2–6.3. Note that here, the rates of convergence for the conditional and unconditional estimators are different. Daraio et al. (2018) develop theory for both the DEA estimators (under convexity of Ψ) and FDH estimators (without imposing convexity). The details require some technical notations but the main idea follows the lines described above in Sections 6.2–6.3. See Daraio et al. (2018) for details and discussion.

6.5 Handling Multiple Sample-Splits

The tests described in Sections 6.2–6.4 require splitting the original sample $\mathcal{X}_n$ into two independent sub-samples $\mathcal{X}_{n_1}$ and $\mathcal{X}_{n_2}$ with $n_1 + n_2 = n$. This introduces some ambiguity and uncertainty due to the particular sample-split, which is typically determined by the choice of the seed for a particular random number generator. Different results (i.e., rejection or failure to reject the null) may result from different splits, and researchers may have difficulty replicating results if they do not have the same random number generator and seed value used to produce the original results. The uncertainty can be largely removed by "integrating out" the randomness of the sample-split by repeatedly splitting the sample a large number of times and appropriately combining information from tests on each of the sample-splits.

To illustrate ideas, suppose that the original sample is split S times. On each split, a test statistic $T_{n,s}$ and a corresponding p-value p_s are computed, with $s = 1, \ldots, S$. Simar and Wilson (2020) explain that it is not easy to exploit the information provided by these S values of test statistic and p-values due to the fact that these quantities are not independent across the different sample-splits. Simar and Wilson (2020) provide a bootstrap algorithm to deal with the dependence across multiple splits, resulting in two tests. The first test is based on the average of the $T_{n,s}$, and the second is based on a Kolmogorov-Smirnov statistic comparing the distribution of the p-values from the S splits to the uniform distribution. Bootstrap methods are used to obtain "global" p-values for the two tests. Unfortunately, the computational burden of the bootstrap method is burdensome and limits the number of splits that can be made in most situations. Nonetheless, Monte Carlo results reported by Simar and Wilson (2020) suggest that the method works well with moderate sample sizes.

Simar and Wilson (2026) provide an improved, fast and easy method for combining results from the multiple sample splits, avoiding the bootstrap. The method is adapted from the multiple testing literature for controlling the false discover rate (e.g., see Benjamini and Yekutieli, 2001). Simar and Wilson (2026, Proposition 4.1) provide a rejection rule that has a rate of type-I error less or equal to the predefined size α. The rule uses only the set of p-value p_s, $s = 1, \ldots, S$. For a test of nominal size α and with β defined as the proportion of splits where $p_s < \alpha^2$, H_0 is rejected if $\beta > \alpha$.

Monte Carlo evidence provided by Simar and Wilson (2026) indicates the method yields tests with good size and power. In some cases, the results are even better than results obtained with a single sample-split or with multiple splits using bootstrap methods. Compared with the bootstrap method of Simar and Wilson (2020) using B bootstrap replications, tests based on the newer method of Simar and Wilson (2026) require computational time equal to approximately B^{-1} times the amount of time consumed by the bootstrap method. Simar and Wilson (2026) provide additional Monte-Carlo evidence suggesting that the number S of sample splits should be 1,000 or more to remove much of the uncertainty arising from the random splitting. Using the new method, one can often use 1,000 or even 10,000 splits with a desktop computer.

7 Dynamic Models

7.1 Malmquist Productivity Indices

Malmquist Productivity Indices (MPIs) based on the work of Malmquist (1953) and Caves et al. (1982) are widely used to measure changes in productivity in dynamic settings (e.g., see Kneip et al., 2021, Appendix A for a discussion of the literature). Unfortunately, many empirical studies present only point estimates of MPIs and geometric means of estimates, with no inference. Until recently, there were no theoretical results allowing empiricists to make inference about MPIs, but this has been remedied by Simar and Wilson (2019) and Kneip et al. (2021, 2026).

In dynamic settings, the attainable set at time t is described by

$$\Psi^t = \{(x, y) \mid x \text{ can produce } y \text{ at time } t\}, \tag{7.1}$$

and its efficient boundary is given by

$$\Psi^{t,\partial} = \left\{(x, y) \mid (x, y) \in \Psi^t,\ (\eta x, \eta^{-1} y) \notin \Psi^t\ \forall\, \eta \in (0, 1)\right\}. \tag{7.2}$$

As discussed by Kneip et al. (2021, 2026), it is advantageous to define the MPI in terms of hyperbolic distances due to the fact that components of some decompositions of the MPI may not be well-defined when the input or output direction is used (see Simar and Wilson, 2023a for an illustration of this problem). Using hyperbolic distance measures, and replacing Ψ with $\mathcal{C}(\Psi)$ in (2.5), the (hyperbolic) MPI measuring the change in productivity for a firm operating at $(x^1, y^1) \in \Psi^1$ at time 1, and at $(x^2, y^2) \in \Psi^2$ at time 2, is defined by

$$\mathcal{M} = \left(\frac{\gamma(x^2, y^2 \mid \mathcal{C}(\Psi^1))}{\gamma(x^1, y^1 \mid \mathcal{C}(\Psi^1))} \times \frac{\gamma(x^2, y^2 \mid \mathcal{C}(\Psi^2))}{\gamma(x^1, y^1 \mid \mathcal{C}(\Psi^2))}\right)^{1/2}. \tag{7.3}$$

It is easy to verify that productivity increases (remains constant, decreases) when $\mathcal{M}$ is greater than (equal to, less than) one.

It is important to note that the distances in (7.3) are measured to from (x^1, y^1) and (x^2, y^2) to the boundaries of $\mathcal{C}(\Psi^1)$ and $\mathcal{C}(\Psi^2)$ instead of to $\Psi^{1,\partial}$ and $\Psi^{2,\partial}$. These distances are well-defined, but they do not measure efficiency unless $\Psi^t = \mathcal{C}(\Psi^t)$ for $t \in \{1,\ 2\}$, i.e., unless CRS holds at times 1 and 2. The measure in (7.3) consists of two ratios. The first measures productivity change for the firm operating at (x^1, y^1) and then (x^2, y^2) using the boundary of $\mathcal{C}(\Psi^1)$ as a benchmark. The second ratio measures productivity change for the same firm using the boundary of $\mathcal{C}(\Psi^2)$ as a benchmark. In addition, it is important to realize that the MPI in (7.3) measures productivity change regardless of whether the production set at times 1 and 2 is convex or whether the frontiers of the production sets exhibit CRS. These are, however, issues for estimation and inference.

7.2 Estimation and Statistical Properties

The MPI defined in (7.3) is unobserved, and hence must be estimated. Suppose n firms are observed at two different times $t \in \{1,\ 2\}$ so that n iid pairs $\{(W_i^1, W_i^2)\}_{i=1}^n$ of observations are available, where $W_i^t = (X_i^t, Y_i^t)$. Then the MPI for a single firm observed operating at $w^t = (x^t, y^t)$ $t = 1,\ 2$ is estimated by

$$\widehat{\mathcal{M}}(w^1, w^2) = \left(\frac{\gamma_\bullet(x^2, y^2 \mid \mathcal{C}(\widehat{\Psi}^1))}{\gamma_\bullet(x^1, y^1 \mid \mathcal{C}(\widehat{\Psi}^1))} \times \frac{\gamma_\bullet(x^2, y^2 \mid \mathcal{C}(\widehat{\Psi}^2))}{\gamma_\bullet(x^1, y^1 \mid \mathcal{C}(\widehat{\Psi}^2))} \right)^{1/2} \tag{7.4}$$

where "$\bullet$" denotes either CFDH or CDEA. CFDH estimators can be used regardless of whether the production sets in the two periods are convex, but use of CDEA estimators requires convexity of the production sets.

Properties of this estimator are based on mild regularity conditions described by Kneip et al. (2021) for the convex case, and by Kneip et al. (2026) for the general case (i.e., where production sets are not necessarily convex). Kneip et al. (2021, 2026) establish that

$$n^\kappa \left(\widehat{\mathcal{M}}(w^1, w^2) - \mathcal{M}(w^1, w^2) \right) \xrightarrow{\mathcal{L}} F_{w^1, w^2} \tag{7.5}$$

as $n \to \infty$, where F_{w^1,w^2} is a non-degenerate distribution depending on the particular estimators used (i.e., CFDH or CDEA) and various features of the DGP (i.e., whether the production sets are convex, smoothness and curvature of the frontier, etc.). Details and discussion are given by Kneip et al. (2021, 2026) The rate κ depends on features of Ψ^1 and Ψ^2 as indicated in Table 1. For example, if CRS does not hold, then $\kappa = 1/(p+q-0.5)$ when CFDH estimators are used, and $\kappa = 2/(p + q + 1)$ if convexity holds and CDEA estimators are used. The theoretical results provided by Kneip et al. (2021, 2026) allow inferences on $\mathcal{M}(w^1, w^2)$ for individual firms to be made using sub-sampling techniques described in Simar and Wilson (2011a).

Empirical studies often report geometric means of MPI estimates for a group of n producers. The geometric mean

$$M_n = \left[\prod_{i=1}^{n} \mathcal{M}(w_i^1, w_i^2) \right]^{1/n} \tag{7.6}$$

can be estimated by

$$\widehat{M}_n = \left[\prod_{i=1}^{n} \widehat{\mathcal{M}}(w_i^1, w_i^2) \right]^{1/n}. \tag{7.7}$$

Kneip et al. (2021, 2026) provide CLTs for both the convex and non-convex cases (respectively) allowing researchers to make inference about average or overall changes in productivity.

Remark 7.1. Unbalanced panels

It is common in empirical studies to discard observations on firms that are only observed in *one of the two periods under consideration in order to work with balanced panels of data. Kneip et al. (2021, 2026) note that this is not necessary. In general, one should never through away data unless it is corrupted by errors that cannot be repaired. While it is true that the estimator* $\widehat{\mathcal{M}}(w^1, w^2)$ *in* (7.4) *can be computed only for firms observed in both periods, the four distance estimates on the right-hand side of* (7.4) *can (and should) be computed using all of the data available in each period. Whenever consistent estimators are used, thawing away data is likely to increase estimation error.*

A number of decompositions of productivity have been considered in the literature. These amount to decomposing (7.3) into sub-indices measuring changes in efficiency, technology, etc. Similar to the MPI defined in (7.3), the various sub-indices resulting from decompositions of (7.3) involve ratios of distances. These sub-indices are typically estimated using the plug-in method, i.e., by replacing unobserved distances with consistent estimators, just as the MPI is estimated by the estimator in (7.4). Statistical properties of estimators of these sub-indices are analogous to properties of estimators of $\mathcal{M}$ discussed above. The relevant theory as well as details and discussion are provided by Simar and Wilson (2019).

Remark 7.2. Other indices
The results of Kneip et al. (2021, 2026) generalize in straightforward ways to Luenberger Productivity Indices using directional distances (see Daraio et al., 2025 for details and discussion). Simar et al. (2025) extend the results to Hicks-Moorsteen Productivity Indices, and Pham et al. (2024) adapt the theory to the case of aggregation of MPIs, where weighted geometric means are constructed while weighting individual MPIs to reflect the economic importance or influence of individual producers.

8 Conclusions and Extensions

For many years, efficiency estimation existed in a vacuum without statistical tools to quantify the uncertainty surrounding estimates of efficiency, productivity, and other items of interest. Today, a rich toolbox of theoretical results and computational methods are available enabling empirical researchers to make practical inference using non-parametric models of production using popular envelopment estimators such as DEA, FDH and their variants. The discussion in this chapter show how crucial the assumptions made on the DGP are—incorrect, unsupported assumptions often lead to inconsistent estimators and futile, invalid inference. Tools now exist with which to test key assumptions within this framework. Many of the techniques described in this chapter can be implemented using a single command in the *FEAR* package for the *R* programming language (see Wilson (2008, 2026 for details). Tasks described above for which there is not single command in *FEAR* can be implemented by writing a few lines of code using the commands in *FEAR*.

Due to space limitations, discussion of some recent and promising extensions has been omitted. the following lines summarize three useful results that should be further developed in the coming years.

First, it is clear that envelopment estimators do not provide useful estimates of marginal products and other features of economic interest. Daraio and Simar (2024) extend previous results from Florens and Simar (2005) by approximating the estimated frontier by smooth estimates, thereby providing local estimates of the partial derivatives of economic interest. These estimates complement exiting methods based on non-parametric envelopment estimators without requiring additional parametric assumptions.

Second, the discussion in Sections 5 and 6.4 shows that conditional frontier models are very useful for incorporating environmental variables in production model. However, these methods require specifying bandwidths to estimate conditional distribution function. Typically, standard LSCV are used to provide these bandwidths. Bădin et al., 2019 observe that LSCV yields bandwidths that may not be optimal for estimating the boundary of the conditional distribution, and data are sparse near this boundary. Florens et al. (2014) and Mastromarco et al. (2025) provide a solution to this problem by proposing an indirect method to estimate the conditional efficiencies by cleaning the inputs and the outputs from the effects of the environmental variables in a first step. Florens et al. (2014) suggest using non-parametric location-scale models for the cleaning step, while Mastromarco et al. (2025) extend the idea to a more general framework by using non-parametric, non-separable models. Monte-Carlo experiments indicates improved performances of the latter method. So far, the method is limited to cases with a one-dimensional response case (e.g., only one input in the input-oriented case).

Finally, the preceding discussion is limited to so-called deterministic frontier models. However, several recent papers introduce stochastic noise in attempts to develop non-parametric stochastic frontier models. Among those that provide theoretical results, Kumbhakar et al. (2007) use a local likelihood method (see also Simar and Zelenyuk, 2011). Kneip et al. (2015a) address identification issues and use pseudo-likelihood techniques and histogram density estimators to approximate the distribution of the efficiencies, while Florens et al. (2020) use Laguerre polynomials to approximate the distribution of efficiencies. Simar et al. (2017) show how a minimal set of local assumptions allows least-squares methods to be used in a non-parametric framework. These approaches are (so far) limited to univariate response variables (e.g., a univariate output in the output orientation). Simar and Wilson (2023b) extend these techniques to fully multivariate settings to provide estimators for almost fully non-parametric stochastic frontier models with multiple inputs and multiple outputs.

References

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

Apon, A. W., L. B. Ngo, M. E. Payne, and P. W. Wilson (2015), Assessing the effect of high performance computing capabilities on academic research output, *Empirical Economics* 48, 283–312.

Aragon, Y., A. Daouia, and C. Thomas-Agnan (2005), Nonparametric frontier estimation: A conditional quantile-based approach, *Econometric Theory* 21, 358–389.

Bădin, L., C. Daraio, and L. Simar (2012), How to measure the impact of environmental factors in a nonparametric production model, *European Journal of Operational Research* 223, 818–833.

— (2014), Explaining inefficiency in nonparametric production models: The state of the art, *Annals of Operations Research* 214, 5–30.

— (2019), A bootstrap approach for bandwidth selection in estimating conditional efficiency measures, *European Journal of Operational Research* 277, 784–797.

Banker, R. D. (1993), Maximum likelihood, consistency and data envelopment analysis: a statistical foundation, *Management Science* 39, 1265–1273.

Banker, R. D., A. Charnes, and W. W. Cooper (1984), Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science* 30, 1078–1092.

Benjamini, Y. and D. Yekutieli (2001), The control of false discovery rate in multiple testing under dependency, *Annals of Statistics* 29, 1165–1188.

Caves, D. W., L. R. Christensen, and W. E. Diewert (1982), The economic theory of index numbers and the measurement of input, output and productivity, *Econometrica* 50, 1393–1414.

Cazals, C., J. P. Florens, and L. Simar (2002), Nonparametric frontier estimation: A robust approach, *Journal of Econometrics* 106, 1–25.

— (2025), Single index models for nonparametric conditional frontiers. LIDAM Discussion paper ISBA 2025/22, UCLouvain.

Chambers, R. G., Y. Chung, and R. Färe (1996), Benefit and distance functions, *Journal of Economic Theory* 70, 407–419.

— (1998), Profit, directional distance functions, and Nerlovian efficiency, *Journal of Optimization Theory and Applications* 98, 351–364.

Charnes, A., W. W. Cooper, and E. Rhodes (1978), Measuring the efficiency of decision making units, *European Journal of Operational Research* 2, 429–444.

Daouia, A., J. P. Florens, and L. Simar (2010), Frontier estimation and extreme value theory, *Bernoulli* 16, 1039–1063.

— (2012), Regularization of non-parametric frontier estimators, *Journal of Econometrics* 168, 285–299.

Daouia, A. and I. Gijbels (2011), Robustness and inference in nonparametric partial frontier modeling, *Journal of Econometrics* 161, 147–165.

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

Daouia, A., L. Simar, and P. W. Wilson (2017), Measuring firm performance using nonparametric quantile-type distances, *Econometric Reviews* 36, 156–181.

Daraio, C., S. Di Leo, and L. Simar (2025), Conical FDH estimators of directional distances and Luenberger productivity indices for general technologies, *European Journal of Operational Research* 24, 93–121.

Daraio, C. and L. Simar (2007a), *Advanced Robust and Nonparametric Methods in Efficiency Analysis*, New York: Springer Science+Business Media, LLC.

— (2007b), Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach, *Journal of Productivity Analysis* 28, 13–32.

— (2014), Directional distances and their robust versions: Computational and testing issues, *European Journal of Operational Research* 237, 358–369.

— (2024), Approximations and inference for nonparametric production frontiers, *Journal of Productivity Analysis* 62, 197–215.

Daraio, C., L. Simar, and P. W. Wilson (2018), Central limit theorems for conditional efficiency measures and tests of the 'separability condition' in non-parametric, two-stage models of production, *The Econometrics Journal* 21, 170–191.

— (2020), Fast and efficient computation of directional distance estimators, *Annals of Operations Research* 288, 805–835.

— (2021), Quality as a latent heterogeneity factor in the efficiency of universities, *Economic Modelling* 99, online publication, `https://doi.org/10.1016/j.econmod.2021.03.004`.

Debreu, G. (1951), The coefficient of resource utilization, *Econometrica* 19, 273–292.

Deprins, D., L. Simar, and H. Tulkens (1984), Measuring labor inefficiency in post offices, in M. M. P. Pestieau and H. Tulkens, eds., *The Performance of Public Enterprises: Concepts and Measurements*, Amsterdam: North-Holland, pp. 243–267.

Färe, R. (1988), *Fundamentals of Production Theory*, Berlin: Springer-Verlag.

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

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

Fève, F., J. P. Florens, and L. Simar (2023), Proportional incremental cost probability functions and their frontiers, *Empirical Economics* 64, 2721–2756.

Florens, J. P. and L. Simar (2005), Parametric approximations of nonparametric frontiers, *Journal of Economietrics* 124, 91–116.

Florens, J. P., L. Simar, and I. Van Keilegom (2014), Frontier estimation in nonparametric location-scale models, *Journal of Econometrics* 178, 456–470.

— (2020), Estimation of the boundary of a variable observed with symmetric error, *Journal of the American Statistical Association* 115, 425–441.

Gijbels, I., E. Mammen, B. U. Park, and L. Simar (1999), On estimation of monotone and concave frontier functions, *Journal of the American Statistical Association* 94, 220–228.

Jeong, S. O. and B. U. Park (2006), Large sample approximation of the distribution for convex-hull estimators of boundaries, *Scandinavian Journal of Statistics* 33, 139–151.

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

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

Kneip, A., B. Park, and L. Simar (1998), A note on the convergence of nonparametric DEA efficiency measures, *Econometric Theory* 14, 783–793.

Kneip, A., L. Simar, and I. Van Keilegom (2015a), Frontier estimation in the presence of measurement error with unknown variance, *Journal of Econometrics* 184, 379–393.

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.

— (2011), A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators, *Computational Economics* 38, 483–515.

— (2015b), When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores, *Econometric Theory* 31, 394–422.

— (2016), Testing hypotheses in nonparametric models of production, *Journal of Business and Economic Statistics* 34, 435–456.

— (2021), Inference in dynamic, nonparametric models of production: Central limit theorems for Malmquist indices, *Econometric Theory* 37, 537–572.

— (2026), Conical FDH estimators for testing returns to scale and making inference about changes in productivity, *Econometric Reviews* Forthcoming. DOI: `https://doi.org/10.1080/07474938.2025.2584132`.

Koopmans, T. C. (1951), An analysis of production as an efficient combination of activities, in T. C. Koopmans, ed., *Activity Analysis of Production and Allocation*, New York: John-Wiley and Sons, Inc., pp. 33–97. Cowles Commission for Research in Economics, Monograph 13.

Korostelev, A., L. Simar, and A. B. Tsybakov (1995a), Efficient estimation of monotone boundaries, *The Annals of Statistics* 23, 476–489.

— (1995b), On estimation of monotone and convex boundaries, *Publications de l'Institut de Statistique de l'Université de Paris XXXIX* 1, 3–18.

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

Kumbhakar, S. C., B. U. Park, L. Simar, and E. G. Tsionas (2007), Nonparametric stochastic frontiers: A local likelihood approach, *Journal of Econometrics* 137, 1–27.

Li, Q., J. Lin, and J. S. Racine (2013), Optimal bandwidth selection for nonparametric conditional distribution and quantile functions, *Journal of Business and Economic Statistics* 31, 57–65.

Malmquist, S. (1953), Index numbers and indifference surfaces, *Trabajos de Estatística y de Investigación Operativa* 4, 209–242.

Mastromarco, C., L. Simar, and I. V. Keilegom (2025), Estimating nonparametric conditional frontiers and efficiencies: A new approach, *The Econometrics Journal* 28, 502–528.

McAllister, P. H. and D. McManus (1993), Resolving the scale efficiency puzzle in banking, *Journal of Banking and Finance* 17, 389–405.

Nguyen, B. H., L. Simar, and V. Zelenyuk (2022), Data sharpening for improving central limit theorem approximations for data envelopment analysis-type efficiency estimators, *European Journal of Operational Research* 303, 1469–1480.

O'Loughlin, C. T. and P. W. Wilson (2021), Benchmarking the performance of U.S. municipalities, *Empirical Economics* 60, 2665–2700.

Park, B. U., S.-O. Jeong, and L. Simar (2010), Asymptotic distribution of conical-hull estimators of directional edges, *Annals of Statistics* 38, 1320–1340.

Park, B. U., L. Simar, and C. Weiner (2000), FDH efficiency scores from a stochastic point of view, *Econometric Theory* 16, 855–877.

Pham, M. D., L. Simar, and V. Zelenyuk (2024), Statistical inference for aggregation of Malmquist productivity indices, *Operations Research* 72, 1615–1628.

Politis, D. N., J. P. Romano, and M. Wolf (2001), On the asymptotic theory of subsampling, *Statistica Sinica* 11, 1105–1124.

Rhodes, E. (1978), *Data Envelopment Analysis and Related Approaches for Measuring the Efficiency of Decision Making Units with an Application to Program Follow Through in U.S. Education*, Ph.D. thesis, Carnegie-Mellon Univesity, School of Urban and Public Affairs, Pittsburgh, PA.

Robinson, P. M. (1988), Root-n-consistent semiparametric regression, *Econometrica* 56, 931–954.

Shephard, R. W. (1970), *Theory of Cost and Production Functions*, Princeton: Princeton University Press.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Simar, L. (1992), Estimating efficiencies from frontier models with panel data: a comparison of parametric, non-parametric and semi-parametric methods with bootstrapping, *Journal of Productivity Analysis* 3, 171–203.

— (1996), Aspects of statistical analysis in DEA-type frontier models, *Journal of Productivity Analysis* 7, 177–185.

— (2003), Detecting outliers in frontier models: A simple approach, *Journal of Productivity Analysis* 20, 391–424.

Simar, L., I. Van Keilegom, and V. Zelenyuk (2017), Nonparametric least squares methods for stochastic frontier models, *Journal of Productivity Analysis* 47, 189–204.

Simar, L. and A. Vanhems (2012), Probabilistic characterization of directional distances and their robust versions, *Journal of Econometrics* 166, 342–354.

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

Simar, L. and P. W. Wilson (1998), Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models, *Management Science* 44, 49–61.

— (2000), A general methodology for bootstrapping in non-parametric frontier models, *Journal of Applied Statistics* 27, 779–802.

— (2001), Testing restrictions in nonparametric efficiency models, *Communications in Statistics* 30, 159–184.

— (2002), Nonparametric tests of returns to scale, *European Journal of Operational Research* 139, 115–132.

— (2007), Estimation and inference in two-stage, semi-parametric models of productive efficiency, *Journal of Econometrics* 136, 31–64.

— (2011a), Inference by the m out of n bootstrap in nonparametric frontier models, *Journal of Productivity Analysis* 36, 33–53.

— (2011b), Two-Stage DEA: Caveat emptor, *Journal of Productivity Analysis* 36, 205–218.

— (2019), Central limit theorems and inference for sources of productivity change measured by nonparametric Malmquist indices, *European Journal of Operational Research* 277, 756–769.

— (2020), Hypothesis testing in nonparametric models of production using multiple sample splits, *Journal of Productivity Analysis* 53, 287–303.

— (2023a), Another look at productivity growth in industrialized countries, *Journal of Productivity Analysis* 60, 257–272.

— (2023b), Nonparametric, stochastic frontier models with multiple inputs and outputs, *Journal of Business and Economic Statistics* 41, 1391–1403.

— (2024), New tools for evaluating the performance of health-care providers using DEA and FDH estimators, in S. Grosskopf, V. Valdmanis, and V. Zelenyuk, eds., *The Cambridge Handbook of Healthcare: Productivity, Efficiency, Effectiveness*, chapter 12, Cambridge, UK: Cambridge University Press, pp. 351–403.

— (2026), A fast method for implementing hypothesis tests with multiple sample splits in nonparametric models of production, *Computational Economics* Forthcoming.

Simar, L. and V. Zelenyuk (2006), On testing equality of two distribution functions of technical efficiency scores estimated via DEA, *Econometric Reviews* 24, 497–522.

— (2007), Statistical inference for aggregates of farrell-type efficiencies, *Journal of Applied Econometrics* 27, 1367–1394.

— (2011), Stochastic FDH/DEA estimators for frontier analysis, *Journal of Productivity Analysis* 36, 1–20.

— (2018), Central limit theorems for aggregate efficiency, *Operations Research* 66, 137–149.

Simar, L., V. Zelenyuk, and S. Zhao (2023), Further improvements of finite sample approximation of central limit theorems for envelopment estimators, *Journal of Productivity Analysis* 59, 189–194.

— (2024), Inference for aggregate efficiency: Theory and guidelines for practitioners, *European Journal of Operational Research* 316, 240–254.

— (2025), Statistical inference for Hicks-Moorsteen productivity indices, *Annals of Operations Research* 351, 1675–1703.

Wheelock, D. C. and P. W. Wilson (2001), New evidence on returns to scale and product mix among U.S. commercial banks, *Journal of Monetary Economics* 47, 653–674.

— (2008), Non-parametric, unconditional quantile estimation for efficiency analysis with an application to Federal Reserve check processing operations, *Journal of Econometrics* 145, 209–225.

— (2012), Do large banks have lower costs? new estimates of returns to scale for U.S. banks, *Journal of Money, Credit, and Banking* 44, 171–199.

— (2018), The evolution of scale-economies in U.S. banking, *Journal of Applied Econometrics* 33, 16–28.

Wikipedia Contributors (2025), Deterministic system — Wikipedia, the free encyclopedia, `https://en.wikipedia.org/w/index.php?title=Deterministic_system&oldid=1276521658`. Online; accessed 15 January 2026.

Wilson, P. W. (1993), Detecting outliers in deterministic nonparametric frontier models with multiple outputs, *Journal of Business and Economic Statistics* 11, 319–323.

— (1995), Detecting influential observations in data envelopment analysis, *Journal of Productivity Analysis* 6, 27–45.

— (2008), *FEAR*: A software package for frontier efficiency analysis with *R*, *Socio-Economic Planning Sciences* 42, 247–254.

— (2011), Asymptotic properties of some non-parametric hyperbolic efficiency estimators, in I. Van Keilegom and P. W. Wilson, eds., *Exploring Research Frontiers in Contemporary Statistics and Econometrics*, Berlin: Springer-Verlag, pp. 115–150.

— (2018), Dimension reduction in nonparametric models of production, *European Journal of Operational Research* 267, 349–367.

— (2021), U.S. banking in the post-crisis era: New results from new methods, in C. Parmeter and R. Sickles, eds., *Advances in Efficiency and Productivity Analysis*, Cham, Switzerland: Springer Nature Switzerland AG, pp. 233–264.

— (2025), A generalized hyperbolic distance function for benchmarking performance: Estimation and inference, *Computational Economics* 65, 3077–3110.

— (2026), Using *FEAR* to analyze efficiency, productivity and model structure. Unpublished working paper, Department of Economics, Clemson University, Clemson, South Carolina 29634 USA.

Wilson, P. W. and S. Zhao (2022), Investigating the performance of Chinese banks over 2007–2014, *Annals of Operations Research* Online publication avaialble at `https://doi.org/10.1007/s10479-022-04925-8`.