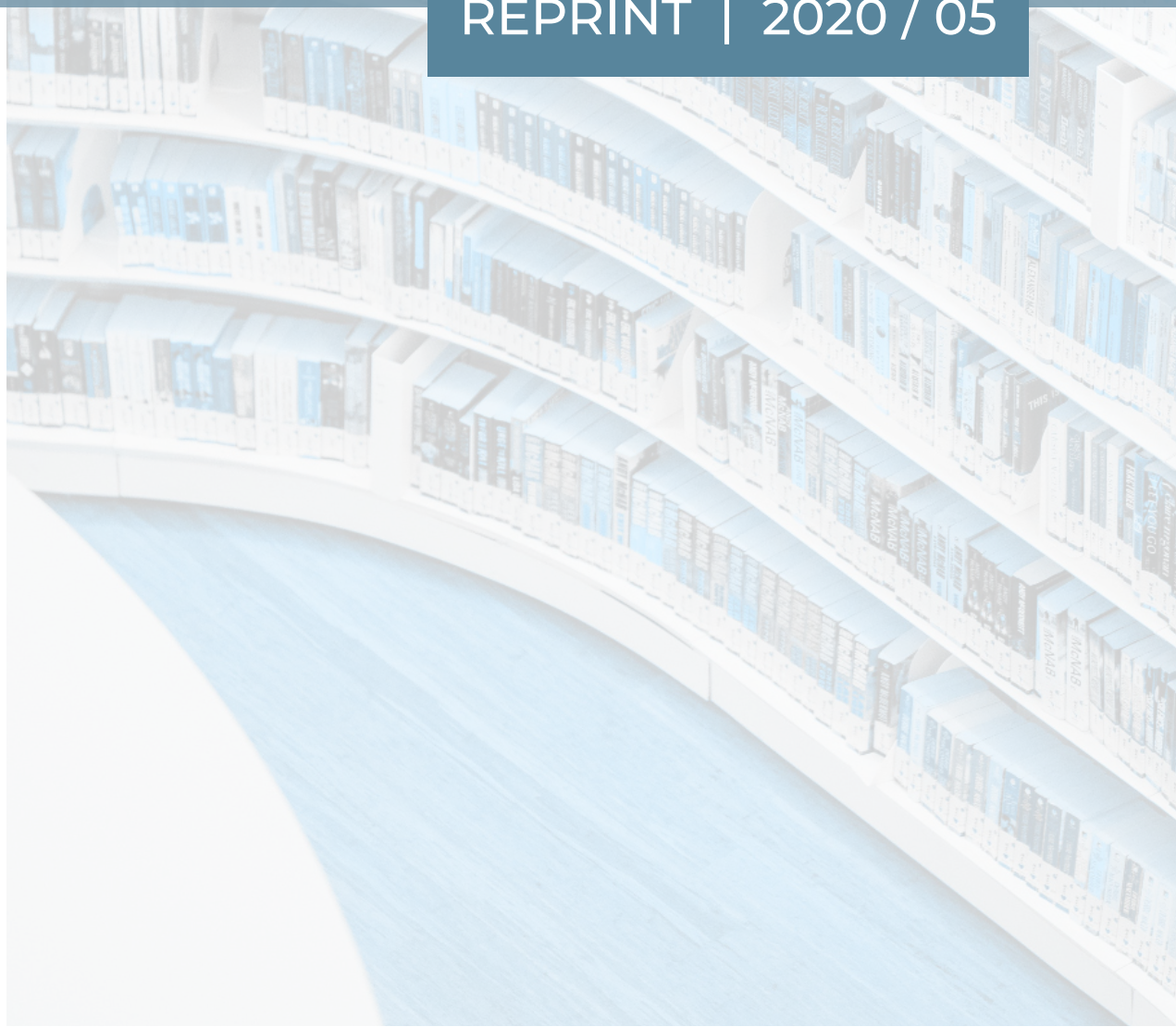
NONPARAMETRIC STATISTICAL ANALYSIS OF PRODUCTION

Mastromarco, C., Simar, L. and P.W. Wilson

REPRINT | 2020 / 05



UCLouvain

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

Nonparametric Statistical Analysis of Production

Camilla Mastromarco Léopold Simar Paul W. Wilson*

March 2018

Abstract

A rich literature on the analysis of efficiency in production has developed since pioneering work of Tjalling Koopmans and George Debreu in the 1950s. This literature includes work by researchers in economics, econometrics, management science, operations research, mathematical statistics and other fields. The focus in this survey is on nonparametric approaches for estimation and inference, with particular emphasis on inference since nothing can be learned from estimation without inference. The statistical problem amounts to estimating the support of a multivariate random variable, subject to some shape constraints, in multiple dimensions. New results that enable inference about mean efficiency as well as tests of convexity, returns to scale, differences in mean efficiency and the separability condition described by Simar and Wilson (2007) are discussed. The well-known curse of dimensionality presents additional challenges, but recent work indicates that reducing dimensionality using eigensystem techniques may improve estimation accuracy. Remaining challenges and open issues are also discussed.

Keywords: Productivity, Efficiency, Data Envelopment Analysis (DEA), Free Disposal Hull (FDH), Nonparametric Frontiers, Boundary Estimation, Panel Data, Extreme Value Theory, Bootstrap

*Mastromarco: Dipartimento di Scienze dell'Economia, Università del Salento, 73100 Lecce, Italy; email camilla.mastromarco@unisalento.it.

Simar: Institut de Statistique, Biostatistique, et Sciences Actuarielles, Université Catholique de Louvain, Voie du Roman Pays 20, B 1348 Louvain-la-Neuve, Belgium; email leopold.simar@uclouvain.be.

Wilson: Department of Economics and Division of Computer Science, School of Computing, 228 Sirrine Hall, Clemson University, Clemson, South Carolina 29634–1309, USA; email pww@clemson.edu.

Research support from IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) is gratefully acknowledged. Any remaining errors are solely our responsibility.

Contents

1 Introduction

Benchmarking performance is a common endeavor. In sports, individual athletes as well as teams of athletes strive to win by performing better than their opponents and to set new records for performance. In medicine, physicians and researchers strive to find treatments that enhance or extend patients' lives better than existing treatments. In education, schools attempt to enhance students' prospects for success at the next level or in the workplace. In manufacturing, firms attempt to convert inputs (e.g., land, labor, capital, materials or energy) into outputs (e.g., goods or services) as "efficiently" (i.e., with as little waste) as possible.

This chapter provides an up-to-date survey of statistical tools and results that are available to applied researchers using nonparametric, deterministic estimators to evaluate producers' performances.[1] Previous surveys (e.g., Simar and Wilson, 2008 and 2013) give summaries of results that allow researchers to estimate and make inference about the efficiency of individual producers. In the review that follows, we describe new results that permit (i) inferences about mean efficiency (both conditional and unconditional), (ii) tests of convexity versus non-convexity of production sets, (iii) tests of constant versus variable returns to scale, (iv) tests of the "separability condition" described by Simar and Wilson (2007) required for second-stage regressions of efficiency estimates on some explanatory variables and (v) dimension reduction to circumvent the slow convergence rates of nonparametric efficiency estimators. We also show how conditional efficiency estimators can be used when panel data are available to model and detect changes over time.

A rich economic theory of efficiency in production has developed from the pioneering work of Debreu (1951) and Koopmans (1951). Farrell (1957) made the first attempt to estimate a measure of efficiency from observed data on production, but the statistical properties of his estimator were developed much later. Researchers working in economics, econometrics, management science, operations research, mathematical statistics, and other fields have

[1] By using "deterministic" to describe these estimators, we follow the language of the literature. The deterministic estimators do not admit a two-sided noise term, in contrast to the literature on efficiency estimation in the context of parametric, stochastic frontier models. Nonetheless, efficiency must be estimated since, as discussed below, it cannot be observed directly. Even if the researcher has available observations on all existing firms in an industry, he should consider that a clever entrepreneur might establish a new firm that out-performs all of the existing firms. Truth cannot be learned from data.

contributed to what has by now become a large literature on efficiency of production.[2]

Measurement of performance (i.e., efficiency) requires a benchmark. A commonly-used benchmark is the efficient production frontier, defined in the relevant input-output space as the locus of the maximal attainable level of outputs corresponding to given levels of inputs.[3] As discussed below, it is also possible to measure efficiency against other benchmarks provided by features that lie "close" to the efficient production frontier. In either case, these benchmarks are unobservable and must be estimated. As such, statistical inference is needed before anything can be learned from data. In turn, a statistical model is needed, as without one inference is not possible as the theoretical results of Bahadur and Savage (1956) make clear.

In the survey that unfolds below, Section 2 presents economic and statistical assumptions that comprise a statistical model in which inference about technical efficiency is possible. Section 3 discusses nonparametric estimators of efficiency. Both cases mentioned above are considered, i.e., where either the efficiency frontier or a feature near the efficient frontier is used as a benchmark. Section 4 introduces "environmental" variables that are neither inputs nor outputs, but which may nonetheless affect production. This requires a new statistical model, which is introduced in Section 4.1 as well as new estimators which are discussed in Section 4.2. Section 5 reviews recent developments providing central limit theorems for mean efficiency when sample means of nonparametric efficiency estimators are used to estimate mean efficiency. Section 6 shows how one can test hypotheses about the structure of the efficient production frontier. Section 7 discusses strategies for increasing accuracy in estimation of efficiency by reducing dimensionality. Section 9 discusses recent developments in dynamic settings, and Section 10 provides conclusions.

[2] On 24 January 2018, a search on Google Scholar using the keywords "efficiency," "production," "inputs" and "outputs" returned approximately 2,590,000 results.

[3] Alternatively, if prices of inputs are available, one can consider a cost frontier defined by the minimal cost of producing various levels of outputs. Or if prices of outputs are available, one can consider a revenue frontier defined by the maximal revenue obtained from using various levels of inputs. If both prices of inputs and outputs are available, one can consider a profit frontier. All of these possibilities relate to allocative efficiency.

2 Theory of Production

2.1 Economic Model

Standard economic theory of the firm (e.g., Koopmans, 1951, Debreu, 1951 or Varian, 1978) describes production in terms of a production set

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

where $x \in \mathbb{R}_+^p$ denote a vector of p input quantities and let $y \in \mathbb{R}_+^q$, denote a vector of q output quantities. The production set consists of the set of physically (or technically) attainable combinations (x, y). For purposes of comparing the performances of producers, the technology (i.e., the efficient boundary of Ψ) given by

$$\Psi^\partial = \{(x, y) \in \Psi \mid (\theta x, \gamma y) \notin \Psi \ \forall \ \theta \in (0, 1) \text{ and } (x, \lambda y) \notin \Psi \ \forall \ \lambda \in (1, \infty)\} \tag{2.2}$$

is a relevant benchmark.[4] Firms that are technically efficient operate in the set Ψ^∂, while those that are technically inefficient operate in the set $\Psi \backslash \Psi^\partial$.[5]

The following economic assumptions are standard in microeconomic theory of the firm (e.g., Shephard, 1970, Färe, 1988).

Assumption 2.1. *Ψ is closed.*

Assumption 2.2. *Both inputs and outputs are freely disposable: if $(x, y) \in \Psi$, then for any (x', y') such that $x' \geq x$ and $y' \leq y$, $(x', y') \in \Psi$.*

Assumption 2.3. *All production requires use of some inputs: $(x, y) \notin \Psi$ if $x = 0$ and $y \geq 0,\ y \neq 0$.*

Closedness of the attainable set Ψ is a mild technical condition, avoiding mathematical problems for infinite production plans. Assumption 2.2 is sometimes called strong disposability and amounts to an assumption of monotonicity of the technology. This property also

[4] In the literature, producers are often referred to as "decision-making units" reflecting the fact that depending on the setting, producers might be firms, government agencies, branches of firms, countries, individuals, or other agents or entities. From this point onward, we will use "firms," which is shorter than "decision-making units," to refer to producers without loss of generality or understanding.

[5] Standard microeconomic theory of the firm (e.g., Varian, 1978) suggests that inefficient firms are driven out of competitive markets. However, the same theory makes no statement about how long this might take. Even in perfectly competitive markets, it is reasonable to believe that inefficient firms exist.

implies the technical possibility of wasting resources, i.e., the possibility of increasing input levels without producing more output, or the possibility of producing less output without reducing levels. Assumption 2.3 means that some amount of input is required to produce any output, i.e., there are no "free lunches."

The Debreu-Farrell (Debreu, 1951 and Farrell, 1957) input measure of *technical* efficiency for a given point $(x, y) \in \mathbb{R}_+^{p+q}$ is given by

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

Note that this measure is defined for some points in $\mathbb{R}_+^{p+q}$ not necessarily in Ψ (i.e., points for which a solution exists in (2.3)). Given an output level y, and an input mix (a direction) given by the vector x, the corresponding efficient level of input is given by

$$x^{\partial}(y) = \theta(x, y \mid \Psi)x, , \tag{2.4}$$

which is the projection of (x, y) onto the efficient boundary Ψ^{∂}, along the ray x orthogonal to the vector y.

In general, for $(x, y) \in \Psi$, $\theta(x, y \mid \Psi)$ gives the feasible proportionate reduction of inputs that a unit located at (x, y) could undertake to become technically efficient. By construction, for all $(x, y) \in \Psi$, $\theta(x, y \mid \Psi) \in (0, 1]$; (x, y) is technically efficient if and only if $\theta(x, y \mid \Psi) = 1$. This measure is the reciprocal of the Shephard (1970) input distance function.

Similarly, in the output direction, the Debreu-Farrell output measure of technical efficiency is given by

$$\lambda(x, y \mid \Psi) = \sup\{\lambda \mid (x, \lambda y) \in \Psi\} \tag{2.5}$$

for $(x, y) \in \mathbb{R}_+^{p+q}$. Analogous to the input-oriented case described above, $\lambda(x, y \mid \Psi)$ gives the feasible proportionate increase in outputs for a unit operating at $(x, y) \in \Psi$ that would achieve technical efficiency. By construction, for all $(x, y) \in \Psi$, $\lambda(x, y \mid \Psi) \in [1, \infty)$ and (x, y) is technically efficient if and only if $\lambda(x, y \mid \Psi) = 1$.

The output efficiency measure $\lambda(x, y \mid \Psi)$ is the reciprocal of the Shephard (1970) output distance function. The efficient level of output, for the input level x and for the direction of the output vector determined by y, is given by

$$y^{\partial}(x) = \lambda(x, y \mid \Psi)y. \tag{2.6}$$

Other distance measures have been proposed in the economic literature, including hyperbolic distance

$$\gamma(x, y \mid \Psi) = \sup\left\{\gamma > 0 \mid (\gamma^{-1}x,\ \gamma y) \in \Psi\right\} \tag{2.7}$$

due to Färe et al. (1985) (see also Färe and Grosskopf, 2004), where input and output quantities are adjusted simultaneously to reach the boundary along an hyperbolic path. Note $\gamma(x, y \mid \Psi) = 1$ if and only if (x, y) belongs to the efficient boundary Ψ^{∂}. Under constant returns to scale (CRS) it is straightforward to show that

$$\gamma(x, y \mid \Psi) = \theta(x, y \mid \Psi)^{-1/2} = \lambda(x, y \mid \Psi)^{1/2}. \tag{2.8}$$

However, no general relationship between the hyperbolic measure and either the input or the output oriented measures hold if the technology does not exhibit CRS everywhere.

Chambers et al. (1998) proposed the directional measure

$$\delta(x, y \mid d_x, d_y, \Psi) = \sup\left\{\delta \mid (x - \delta d_x, y + \delta d_y) \in \Psi\right\}, \tag{2.9}$$

which measures the distance from a point (x, y) to the frontier in the given direction $d = (-d_x, d_y)$, where $d_x \in \mathbb{R}^p_+$ and $d_y \in \mathbb{R}^q_+$. This measure is flexible in the sense that some values of the direction vector can be set to zero. A value $\delta(x, y \mid d_x, d_y, \Psi) = 0$ indicates an efficient point lying on the boundary of Ψ. Note that as a special case, the Farrell-Debreu radial distances can be recovered; e.g. if $d = (-x, 0)$ then $\delta(x, y \mid d_x, d_y, \Psi) = 1 - \theta(x, y \mid \Psi)^{-1}$ or if $d = (0, y)$ then $\delta(x, y \mid d_x, d_y, \Psi) = \lambda(x, y \mid \Psi) - 1$. Another interesting feature is that directional distances are additive measures, hence they permit negative values of x and y (e.g., in finance, an output y may be the return of a fund, which can be, and often is, negative).[6] Many choices of the direction vector are possible (e.g., a common one for all firms, or a specific direction for each firm; see Färe et al., 2008 for discussion), although care should be taken to ensure that the chosen direction vector maintains invariance with respect to units of measurement for input and output quantities.

All of the efficiency measures introduced above characterize the efficient boundary by measuring distance from a known, fixed point (x, y) to the unobserved boundary Ψ^{∂}. The

[6] The measure in (2.9) differs from the "additive" measure $\zeta(x, y \mid \Psi) = \sup\{(i_p' d_x + i_q' d_y \mid (x - d_x, y + d_y) \in \Psi\}$ estimated by Charnes et al. (1985), where $i_p,\ i_q$ denote $(p \times 1)$ and $(q \times 1)$ vectors of ones. Charnes et al. (1985) present only an estimator, and do not define the object that is estimated. Moreover, the additive measure is not in general invariant to units of measurement.

only difference among the measures in (2.3)–(2.9) is in the direction in which distance is measured. Consequently, the remainder of this chapter will focus on the output direction, with references given as appropriate for details on the other directions.

2.2 Statistical Model

The economic model described above introduces a number of concepts, in particular the production set Ψ and the various measures of efficiency, that are unobservable and must be estimated from data, i.e., from a sample $\mathcal{S}_n = \{(X_i, Y_i)\}_{i=1}^n$ of n observed input-output pairs. Nonparametric estimators of the efficiency estimators introduced in Section 2.1 are based either on the free-disposal hull (FDH), the convex hull or the conical hull of the sample observations. These estimators are referred to below as FDH, VRS-DEA (where "VRS" denotes variable returns to scale and "DEA" denotes data envelopment analysis) and CRS-DEA estimators, and are discussed explicitly later in Section 3.

Before describing the estimators it is important to note that the theoretical results of Bahadur and Savage (1956) make clear the need for a statistical model. Such a model is defined here through the assumptions that follow, in conjunction with the assumptions discussed above. The assumptions given here correspond to those in Kneip et al. (2015b).[7] The assumptions are stronger than those used by Kneip et al. (1998), Kneip et al. (2008), Park et al. (2000) and Park et al. (2010) to establish consistency, rates of convergence and limiting distributions for FDH and DEA estimators, but are needed to establish results on moments of the estimators.

Assumption 2.4. *(i) The sample observations $\mathcal{S}_n = \{(X_i, Y_i)\}_{i=1}^n$ are realizations of independent, identically distributed (iid) random variables (X, Y) with joint density f and compact support $\mathcal{D} \subset \Psi$; and (ii) f is continuously differentiable on $\mathcal{D}$.*

The compact set $\mathcal{D}$ is introduced in Assumption 2.4 for technical reasons and is used in proofs of consistency of DEA and FDH estimators; essentially, the assumption rules out use of infinite quantities of one or more inputs.

Assumption 2.5. *(i) $\mathcal{D}^* := \{(x, \lambda(x, y \mid \Psi)y) \mid (x, y) \in \mathcal{D}\} \subset \mathcal{D}$; (ii) $\mathcal{D}^*$ is compact; and (iii) $f(x, \lambda(x, y \mid \Psi)y) > 0$ for all $(x, y) \in \mathcal{D}$.*

[7] Kneip et al. (2015b) work in the input orientation. Here, assumptions are stated for the output orientation where appropriate or relevant.

Assumption 2.5(i) ensures that the projection of any firm in $\mathcal{D}$ onto the frontier in the output direction also is contained in $\mathcal{D}$, and part (ii) means that the set of such projections is both closed and bounded. Part (iii) of the assumption ensures that the density $f(x, y)$ is positive along the frontier where firms in $\mathcal{D}$ are projected, and together with Assumption 2.4 this precludes a probability mass along the frontier. Consequently, observation of a firm with no inefficiency is an event with zero measure; i.e., any given firm almost surely operates in the interior of Ψ.

Assumption 2.6. *$\lambda(x, y \mid \Psi)$ is three times continuously differentiable on $\mathcal{D}$.*

Assumption 2.6 amounts to an assumption of smoothness of the frontier. Kneip et al. (2015b) require only two-times differentiability to establish the limiting distribution of the VRS-DEA estimator, but more smoothness is required to establish moments of the estimator. In the case of the FDH estimator, Assumption 2.6 can be replaced by the following.

Assumption 2.7. *(i) $\lambda(x, y \mid \Psi)$ is twice continuously differentiable on $\mathcal{D}$; and (ii) all the first-order partial derivatives of $\lambda(x, y \mid \Psi)$ with respect to x and y are nonzero at any point $(x, y) \in \mathcal{D}$.*

Note that the free disposability assumed in Assumption 2.2 implies that $\lambda(x, y \mid \Psi)$ is monotone, increasing in x and monotone, decreasing in y. Assumption 2.7 additionally requires that the frontier is strictly monotone and does not possess constant segments (which would be the case, for example, if outputs are discrete as opposed to continuous, as in the case of ships produced by shipyards). Finally, part (i) of Assumption 2.7 is weaker than Assumption 2.6; here the frontier is required to be smooth, but not as smooth as required by Assumption 2.6.[8]

For the VRS-DEA estimator, the following assumption is needed.

Assumption 2.8. *$\mathcal{D}$ is* almost strictly convex*; i.e., for any $(x, y), (\widetilde{x}, \widetilde{y}) \in \mathcal{D}$ with $(\frac{x}{\|x\|}, y) \neq (\frac{\widetilde{x}}{\|\widetilde{x}\|}, \widetilde{y})$, the set $\{(x^*, y^*) \mid (x^*, y^*) = (x, y) + \alpha((\widetilde{x}, \widetilde{y}) - (x, y)) \text{ for some } 0 < \alpha < 1\}$ is a subset of the* interior *of $\mathcal{D}$.*

Assumption 2.8 replaces the assumption of strict convexity of Ψ used in Kneip et al. (2008) to establish the limiting distribution of the VRS-DEA estimator. Together with

[8] Assumption 2.7 is slightly stronger, but much simpler than assumptions AII–AIII in Park et al. (2000).

Assumption 2.5, Assumption 2.8 ensures that the frontier Ψ^{∂} is convex in the region where observations are projected onto Ψ^{∂} by $\lambda(x, y \mid \Psi)$. Note, however, that convexity may be a dubious assumption in some situations. For example, see Bogetoft (1996), Bogetoft et al. (2000), Briec et al. (2004)) and Apon et al. (2015). Convexity is not needed when FDH estimators are used, and in Section 6.1 a statistical test of the assumption is discussed.

For the case of the CRS-DEA estimator, Assumption 2.8 must be replaced by the following condition.

Assumption 2.9. *(i) For any* $(x, y) \in \Psi$ *and any* $a \in [0, \infty)$, $(ax, ay) \in \Psi$*; (ii) the support* $\mathcal{D} \subset \Psi$ *of* f *is such that for any* (x, y), $(\widetilde{x}, \widetilde{y}) \in \mathcal{D}$ *with* $(\frac{x}{\|x\|}, \frac{y}{\|y\|}) \neq (\frac{\widetilde{x}}{\|\widetilde{x}\|}, \frac{\widetilde{y}}{\|\widetilde{y}\|})$, *the set* $\{(x^*, y^*) \mid (x^*, y^*) = (x, y) + \alpha((\widetilde{x}, \widetilde{y}) - (x, y))$ *for some* $0 < \alpha < 1\}$ *is a subset of the* interior *of* $\mathcal{D}$*; and (iii)* $(x, y) \notin \mathcal{D}$ *for any* $(x, y) \in \mathbb{R}^p_+ \times \mathbb{R}^q$ *with* $y^1 = 0$*, where* y^1 *denotes the first element of the vector* y*.*

The conditions on the structure of Ψ (and $\mathcal{D}$) given in Assumptions 2.8 and 2.9 are incompatible. It is not possible that both assumptions hold simultaneously. Assumption 2.9(i) implies that the technology Ψ^{∂} is characterized by globally CRS. Under Assumption 2.9(i), Ψ is equivalent to its convex cone, denoted by $\mathcal{V}(\Psi)$. Otherwise, $\Psi \subset \mathcal{V}(\Psi)$ and (provided Assumptions 2.5 and 2.8 hold) Ψ^{∂} is said to be characterized by VRS.

Assumptions 2.4–2.6 are similar to assumptions needed by Kneip et al. (2008) to establish the limiting distribution of the VRS-DEA estimator, except that there and as noted above, the efficiency measure is only required to be twice continuously differentiable. Here, the addition of Assumption 2.8 (or Assumption 2.9 in the CRS case) and the additional smoothness of $\lambda(x, y \mid \Psi)$ in Assumption 2.6 are needed to establish results beyond those obtained in Kneip et al. (2008).

2.3 An Alternative Probabilistic Framework

The description of the DGP in Section 2.2 is traditional. However, the DGP can also be described in terms that allow a probabilistic interpretation of the Debreu-Farrell efficiency scores, providing a new way of describing the nonparametric FDH and DEA estimators. This new formulation is useful for introducing extensions of the FDH and DEA estimators described above, and for linking frontier estimation to extreme value theory as explained by

Daouia et al. (2010). The presentation here draws on that of Daraio and Simar (2005), who extend the ideas of Cazals et al. (2002).

The stochastic part of the DGP introduced in Section 2.2 through the probability density function $f(x, y)$ is completely described by the distribution function

$$H_{XY}(x, y) = \Pr(X \leq x, Y \geq y). \tag{2.10}$$

This is not a standard distribution function since the cumulative form is used for the inputs x while the survival form is used for the outputs y. However, $H_{XY}(x, y)$ is well-defined and gives the probability that a unit operating at input, output levels (x, y) is *dominated,* i.e., that another unit produces at least as much output while using no more of any input than the unit operating at (x, y). The distribution function is monotone, non-decreasing in x and monotone non-increasing in y. In addition, the support of the distribution function $H_{XY}(\cdot, \cdot)$ is the attainable set Ψ; i.e.,

$$H_{XY}(x, y) = 0 \; \forall \; (x, y) \notin \Psi. \tag{2.11}$$

The joint probability $H_{XY}(x, y)$ can be decomposed using Bayes' rule to obtain

$$H_{XY}(x, y) = \underbrace{\Pr(X \leq x \mid Y \geq y)}_{=F_{X|Y}(x|y)} \underbrace{\Pr(Y \geq y)}_{=S_Y(y)} \tag{2.12}$$

$$= \underbrace{\Pr(Y \geq y \mid X \leq x)}_{=S_{Y|X}(y|x)} \underbrace{\Pr(X \leq x)}_{=F_X(x)}, \tag{2.13}$$

where $S_Y(y) = \Pr(Y \geq y)$ denotes the survivor function of Y, $S_{Y|X}(y \mid x) = \Pr(Y \geq y \mid X \leq x)$ denotes the conditional survivor function of Y, and the conditional distribution and survivor functions are assumed to exist whenever used (i.e., when needed, $S_Y(y) > 0$ and $F_X(x) > 0$). The frontier Ψ^{∂} can be defined in terms of the conditional distributions defined by (2.12) and (2.13) since the support of $H(x, y)$ is the attainable set. This permits definition of some alternative concepts of efficiency.

For the output-oriented case, assuming $F_X(x) > 0$, define

$$\begin{aligned} \widetilde{\lambda}(x, y \mid H_{XY}) &= \sup\{\lambda \mid S_{Y|X}(\lambda y \mid x) > 0\} \\ &= \sup\{\lambda \mid H_{XY}(x, \lambda y) > 0\}. \end{aligned} \tag{2.14}$$

The output efficiency score $\widetilde{\lambda}(x, y \mid H_{XY})$ gives the proportionate increase in outputs required for the same unit to have zero probability of being dominated by a randomly chosen unit,

holding input levels fixed. Note that in a multivariate framework, the radial nature of the Debreu-Farrell measures is preserved. Similar definitions are possible in the input, hyperbolic and directional orientations; see Simar and Wilson (2013) for details and references.

From the properties of the distribution function $H_{XY}(x, y)$, it is clear that the new efficiency score defined in (2.14) has some sensible properties. First, $\widetilde{\lambda}(x, y \mid H_{XY})$ is monotone, non-decreasing in x and monotone, non-increasing in y. Second, and most importantly, under Assumption 2.2 it is trivial to show that $\widetilde{\lambda}(x, y \mid H_{XY}) = \lambda(x, y \mid \Psi)$. Therefore, whenever Assumption 2.2 holds, the probabilistic formulation presented here provides an alternative characterization of the traditional Debreu-Farrell efficiency scores.

2.4 Partial Frontiers

As will be seen below in Section 3, estimators of the various efficiency measures introduced above are sensitive to outliers and are plagued by slow convergence rates that depend on $(p + q)$, i.e., the number of inputs and outputs. One approach to avoid these problems is to measure efficiency relative to some benchmark other than the *full frontier* Ψ^{∂}. The notion of *partial frontiers* described below provides useful, alternative benchmarks against which producers' performances can be measured. The main idea is to replace the full frontier with a model feature that lies "close" to the full frontier. Using estimates of distance to a partial frontier, one can rank firms in terms of their efficiency just as one can do when measuring distance to the full frontier.

The density of inputs and outputs introduced in Assumption 2.4 and the corresponding distribution function in (2.10) permit definition of other benchmarks (in addition to Ψ^{∂}) for evaluating producers' performances. Two classes of partial frontiers have been proposed: (i) order-m frontiers, where m can be viewed as a trimming parameter, and (ii) order-α quantile frontiers, analogous to traditional quantile functions but adapted to the frontier problem. As will be seen below in Sections 3.3 and 3.4, estimation based on the concept of partial frontiers offers some advantages over estimation based on the full frontier Ψ^{∂}. In particular, partial frontiers are less sensitive to outliers in the data, avoided and root-n convergence rates are achieved by estimators based on partial frontier concepts.

Suppose input levels x in the interior of the support of X is given, and consider m iid random variables Y_i, $i = 1, \ldots, m$ drawn from the conditional q-variate distribution function

$F_{Y|X}(y \mid x) = 1 - S_{Y|X}(y \mid x) = \Pr(Y \leq y \mid X \leq x)$. Define the random set

$$\Psi_m(x) = \bigcup_{i=1}^{m} \left\{ (\widetilde{x}, \widetilde{y}) \in \mathbb{R}_+^{p+q} \mid \widetilde{x} \leq x, Y_i \leq \widetilde{y} \right\}. \tag{2.15}$$

Then the *random* set $\Psi_m(y)$ is the free disposal hull of m randomly-chosen firms that use no more than x levels of the p inputs. For any y and the given x, the Debreu-Farrell output efficiency score relative to the set $\Psi_m(x)$ is simply

$$\lambda(x, y \mid \Psi_m(x)) = \sup\{\lambda \mid (x, \lambda y) \in \Psi_m(x)\} \tag{2.16}$$

after substitution of $\Psi_m(x)$ for Ψ in (2.5). Of course, the efficiency score $\lambda(x, y \mid \Psi_m(x))$ is random, since the set $\Psi_m(x)$ is random. For a given realization of the m values Y_i, a realization of $\lambda(x, y \mid \Psi_m(x))$ is determined by

$$\lambda(x, y \mid \Psi_m(x)) = \max_{i=1,\ldots,m} \left\{ \min_{j=1,\ldots,p} \left(\frac{Y_i^j}{y^j} \right) \right\} \tag{2.17}$$

where superscript j indexes elements of the q-vectors y and Y_i. Then for any $y \in \mathbb{R}_+^q$, the (expected) order-m output efficiency measure is defined for all x in the interior of the support of X by

$$\begin{aligned} \lambda_m(x, y \mid H_{XY}) =& E(\lambda_m(x, y \mid \Psi_m(x)) \mid X \leq x) \\ =& \int_0^\infty \left[1 - (1 - S_{Y|X}(uy \mid x))^m \right] du \\ =& \lambda(x, y \mid \Psi) - \int_0^{\lambda(x,y|\Psi)} (1 - S_{Y|X}(uy \mid x))^m \, du \end{aligned} \tag{2.18}$$

provided the expectation exists. Clearly,

$$\lim_{m \to \infty} \lambda_m(x, y \mid H_{XY}(x, y)) = \lambda(x, y \mid \Psi). \tag{2.19}$$

The order-m output efficiency score provides a benchmark for the unit operating at (x, y) relative to the expected maximum output among m peers drawn randomly from the population of units that use no more than x levels of inputs. This efficiency measure in turn defines an order-m output efficient frontier. For any $(x, y) \in \Psi$, the expected maximum level of outputs of order-m for a unit using input level x and for an output mix determined by the vector y is given by

$$y_m^\partial(x) = \lambda(x, y \mid \Psi_m(x))\, y. \tag{2.20}$$

The expected output efficient frontier of order m (i.e., the output order-m efficient frontier) is defined by

$$\Psi_m^{\partial \text{out}} = \{(\widetilde{x}, \widetilde{y}) \mid \widetilde{x} = x,\ \widetilde{y} = y\lambda_m(x, y),\ (x, y) \in \Psi\}\,. \tag{2.21}$$

By contrast, recall that the full frontier Ψ^∂ is defined (at input level x) by (2.6).

Extension to the input oriented case is straightforward; see Cazals et al. (2002) for details. Extension to hyperbolic and directional distances is somewhat more complicated due to the nature of the order-m concept in the multivariate framework, and requires some additional work. Results are given by Wilson (2011) for the hyperbolic case, and by Simar and Vanhems (2012) for directional cases.

The central idea behind order-m estimation is to substitute a feature "close" to the frontier Ψ^∂ for the frontier itself. Introduction of the distribution function $H_{XY}(x, y)$ invites a similar, alternative approach based on quantiles. As noted above, the quantity m in order-m frontier estimation serves as a trimming parameter which determines the percentage of points that will lie above the order-m frontier. The idea underlying order-α quantile frontiers is to reverse this causation, and choose the proportion of the data lying above the partial frontier directly.

Quantile estimation in regression contexts is an old idea. In the framework of production frontiers, using the probabilistic formulation of the DGP developed in Section 2.3, it is straightforward to adapt the order-m ideas to order-α quantile estimation. These ideas are developed for the univariate case in the input and output orientations by Aragon et al. (2005) and extended to the multivariate setting by Daouia and Simar (2007). Wheelock and Wilson (2008) extended the ideas to the hyperbolic orientation, and Simar and Vanhems (2012) extended the ideas to directional measures.

Consider again the output oriented case. For all x such that $F_X(x) > 0$ and for $\alpha \in (0, 1]$, the output α-quantile efficiency score for the unit operating at $(x, y) \in \Psi$ is defined by

$$\lambda_\alpha(x, y \mid H_{XY}) = \sup\{\lambda \mid S_{Y|X}(\lambda y \mid x) > 1 - \alpha\}. \tag{2.22}$$

To illustrate this measure, suppose that $\lambda_\alpha(x, y) = 1$. Then the unit operating at $(x, y) \in \Psi$ is said to be output efficient at the level $(\alpha \times 100)$-percent, meaning that the unit is dominated with probability $(1 - \alpha)$ by firms using *no more than* x level of inputs. More generally, if $\lambda_\alpha(x, y)(<, >)1$, the firm at (x, y) can (decrease, increase) its output to $\lambda_\alpha(x, y)y$ to become

output-efficient at level $(\alpha \times 100)$-percent, i.e., to be dominated by firms using weakly less input (than the level x) with probability $(1-\alpha)$.

The concept of order-α output efficiency allows definition of the corresponding efficient frontier at the level $(\alpha \times 100)$-percent. For a given (x, y), the order-α output efficiency level of outputs is given by

$$y_\alpha^\partial(x) = \lambda_\alpha(x, y \mid H_{XY})y. \tag{2.23}$$

By construction, a unit operating at the point $(x, y_\alpha^\partial(x)) \in \Psi$ has a probability $H_{XY}(x, y_\alpha^\partial(x)) = (1-\alpha)F_X(x) \leq 1-\alpha$ of being dominated. Analogous to $\Psi_m^{\partial\text{out}}$, $y_\alpha^\partial(x)$ can be evaluated for all possible x to trace out an order-α output frontier, denoted by $\Psi_\alpha^{\partial\text{out}}$.

It is straightforward to show that $\lambda_\alpha(x, y \mid H_{XY})$ converges monotonically to the Debreu-Farrell output efficiency measure; i.e.,

$$\lim_{\alpha \uparrow 1} \lambda_\alpha(x, y \mid H_{XY}) = \lambda(x, y \mid \Psi) \tag{2.24}$$

where "$\uparrow$" denotes monotonic convergence from below. Moreover, for all $(x, y) \in \Psi$, $(x, y) \notin \Psi^\partial$, there exists an $\alpha \in (0, 1]$ such that $\lambda_\alpha(x, y \mid H_{XY}) = 1$, where $\alpha = 1 - S_{Y|X}(y \mid x)$.

From the preceding discussion it is clear that both the output order-m and output order-α partial frontiers must lie weakly below the full frontier. Consequently, both $\lambda_m(x, y \mid H_{XY}(x, y))$ and $\lambda_\alpha(x, y \mid H_{XY}(x, y))$ must be weakly less than $\lambda(x, y \mid \Psi)$. Conceivably, $\lambda_m(x, y \mid H_{XY}(x, y))$ or $\lambda_\alpha(x, y \mid H_{XY}(x, y))$ could be less than one, indicating that (x, y) lies above the corresponding partial frontier. This is natural since by construction, the partial frontiers (provided $\alpha < 1$ or m is finite) lie in the interior of Ψ. As noted above, partial frontiers provide alternative benchmarks against which efficiency can be measured, with the advantage that the partial frontiers can be estimated with root-n convergence rate and with less sensitivity to outliers than estimates based on full frontiers.

Similar ideas have been developed for other directions. For the input orientation, see Daouia and Simar (2007). Wheelock and Wilson (2008) extend the ideas of Daouia and Simar (2007) to the hyperbolic direction, while Simar and Vanhems (2012) extend these ideas to the directional case.

It is important to note that all of the concepts discussed so far describe unobservable features of the statistical model. Consequently, efficiency must be estimated, and then

inference must be made before anything can be learned about efficiency. The preceding discussion makes clear that more than one benchmark is available by which efficiency can be measured, e.g., the full frontier Ψ^∂, or the order-m or order-α partial frontiers. In any case, these are not observed and must be estimated. The next section discusses various nonparametric estimators of efficiency.

3 Nonparametric Estimation of Efficiency

3.1 Free Disposal Hull Estimators

The FDH estimator

$$\widehat{\Psi}_{\text{FDH}} = \bigcup_{(X_i,Y_i)\in\mathcal{S}_n} \{(x,y) \in \mathbb{R}^{p+q}_+ \mid x \geq X_i,\ y \leq Y_i\}, \tag{3.1}$$

of Ψ proposed by Deprins et al. (1984) relies on free disposability assumed in Assumption 2.2, and does not require convexity of Ψ. The FDH estimator defined in (3.1) is simply the free disposal hull of the observed sample $\mathcal{S}_n$, and amounts to the union of n southeast-orthants with vertices (X_i, Y_i), where n is the number of input-output pairs in $\mathcal{S}_n$.

A nonparametric estimator of output efficiency for a given point $(x, y) \in \mathbb{R}^{p+q}_+$ is obtained by replacing the true production set Ψ in (2.5) with the estimator $\widehat{\Psi}_{\text{FDH}}$, yielding

$$\widehat{\lambda}_{\text{FDH}}(x, y \mid \mathcal{S}_n) = \sup\{\lambda \mid (x, \lambda y) \in \widehat{\Psi}_{\text{FDH}}\}. \tag{3.2}$$

This can be computed quickly and easily by first identifying the set

$$D(x,y) = \{i \mid (X_i, Y_i) \in \mathcal{S}_n,\ X_i \leq x,\ Y_i \geq y\} \tag{3.3}$$

of indices of observations (X_i, Y_i) that dominate (x, y) and then computing

$$\widehat{\lambda}_{\text{FDH}}(x, y \mid \mathcal{S}_n) = \max_{i\in D(x,y)} \min_{j=1,\ldots,q} \left(\frac{Y_i^j}{y^j} \right). \tag{3.4}$$

Determining the set $D(x, y)$ requires only some partial sorting and some simple logical comparisons. Consequently, the estimator $\widehat{\lambda}_{\text{FDH}}(x, y \mid \mathcal{S}_n)$ can be computed quickly and easily. Efficient output levels for given input levels x and a given output mix (direction) described by the vector y are estimated by

$$\widehat{y}^\partial(x) = \widehat{\lambda}_{\text{FDH}}(x, y \mid \mathcal{S}_n) y. \tag{3.5}$$

By construction, $\widehat{\Psi}_{\text{FDH}} \subseteq \Psi$, and so $\widehat{\Psi}_{\text{FDH}}$ is a downward, inward-biased estimator of Ψ, and $\widehat{y}^{\partial}(x)$ is an downward-biased estimator of $y^{\partial}(x)$ defined in (2.6). Consequently, the level of technical efficiency is under-stated by $\widehat{\lambda}_{\text{FDH}}(x, y \mid \mathcal{S}_n)$.

The plug-in principle used above can be used to define an FDH estimator of the input-oriented efficiency measure in (2.3). Wilson (2011) extends this to the hyperbolic case, and Simar and Vanhems (2012) extend the idea to the directional case.

Results from Park et al. (2000) and Daouia et al. (2017) for the input-oriented case extend to the output oriented case described here. In particular,

$$n^{1/(p+q)} \left(\widehat{\lambda}_{\text{FDH}}(x, y \mid \mathcal{S}_n) - \lambda(x, y \mid \Psi) \right) \xrightarrow{\mathcal{L}} \text{Weibull}(\mu_{xy},\ p+q) \tag{3.6}$$

where μ_{xy} is a constant that depends on the DGP; see Park et al. (2000) for details. Wilson (2011) and Simar and Vanhems (2012) establish similar results for the hyperbolic and directional cases. Regardless of the direction, the various FDH efficiency estimators converge at rate $n^{1/(p+q)}$. For $(p+q) > 2$, the FDH estimators converge slower than the parametric root-n rate.

3.2 Data Envelopment Analysis Estimators

Although DEA estimators were first used by Farrell (1957) to measure technical efficiency for a set of observed firms, the idea did not gain widespread notice until Charnes et al. (1978) appeared. Charnes et al. used the convex cone (rather than the convex hull) of $\widehat{\Psi}_{\text{FDH}}$ to estimate Ψ, which would be appropriate only if returns to scale are everywhere constant. Later, Banker et al. (1984) used the convex hull of $\widehat{\Psi}_{\text{FDH}}$ to estimate Ψ, thus allowing variable returns to scale.[9] Here, "DEA" refers to both of these approaches, as well as other approaches that involve definition of a convex set enveloping the FDH estimator $\widehat{\Psi}_{\text{FDH}}$ to estimate Ψ.

The most general DEA estimator of the attainable set Ψ is simply the convex hull of $\widehat{\Psi}_{\text{FDH}}$, i.e.,

$$\widehat{\Psi}_{\text{VRS}} = \{(x, y) \in \mathbb{R}^{p+q} \mid y \leq \boldsymbol{Y}\omega,\ x \geq \boldsymbol{X}\omega,\ i_n'\omega = 1,\ \omega \in \mathbb{R}^n_+\} \tag{3.7}$$

[9] Confusingly, both Charnes et al. (1978) and Banker et al. (1984) refer to their estimators as "models" instead of "estimators." The approach of both is a-statistical, but the careful reader will remember that truth cannot be learned from data.

where $\boldsymbol{X} = \begin{bmatrix} X_1 & \ldots, X_n \end{bmatrix}$ and $\boldsymbol{Y} = \begin{bmatrix} Y_1 & \ldots, Y_n \end{bmatrix}$ are $(p \times n)$ and $(q \times n)$ matrices (respectively) whose columns are the input-output combinations in $\mathcal{S}_n$, ω is an $(n \times 1)$ vector of weights, and i_n is an $(n \times 1)$ vector of ones. Alternatively, the conical hull of the FDH estimator, $\widehat{\Psi}_{\text{CRS}}$, used by Charnes et al. (1978), is obtained by dropping the constraint $i_n\omega = 1$ in (3.7); i.e.,

$$\widehat{\Psi}_{\text{CRS}} = \{(x, y) \in \mathbb{R}^{p+q} \mid y \leq \boldsymbol{Y}\omega,\ x \geq \boldsymbol{X}\omega,\ \omega \in \mathbb{R}^n_+\}. \tag{3.8}$$

As with the FDH estimators, DEA estimators of the efficiency scores $\theta(x, y \mid \Psi)$, $\lambda(x, y \mid \Psi)$ $\gamma(x, y \mid \Psi)$, and $\delta(x, y \mid \boldsymbol{u}, \boldsymbol{v}, \Psi)$ defined in (2.3), (2.5), (2.7), and (2.9) can be obtained using the plug-in method by replacing the true, but unknown, production set Ψ with one of the estimators $\widehat{\Psi}_{\text{VRS}}$ or $\widehat{\Psi}_{\text{CRS}}$. In the output orientation, replacing Ψ with $\widehat{\Psi}_{\text{VRS}}$ in (2.5) yields

$$\widehat{\lambda}_{\text{VRS}}(x, y \mid \mathcal{S}_n) = \sup\left\{\lambda \mid (x, \lambda y) \in \widehat{\Psi}_{\text{VRS}}\right\}, \tag{3.9}$$

which can be computed by solving the linear program

$$\widehat{\lambda}_{\text{VRS}}(x, y \mid \mathcal{S}_n) = \left\{\lambda \mid \lambda y \leq \boldsymbol{Y}\omega,\ x \geq \boldsymbol{X}\omega,\ i_n'\omega = 1,\ \omega \in \mathbb{R}^n_+\right\}. \tag{3.10}$$

The CRS estimator $\widehat{\lambda}_{\text{CRS}}(x, y \mid \mathcal{S}_n)$ is obtained by dropping the constraint $i_n'\omega = 1$ on the right-hand side of (3.10). Technically efficient levels of outputs can be estimated by plugging either the VRS or CRS version of the DEA efficiency estimator into (2.6). For example, under VRS, in the output orientation the technically efficient level of inputs for a given level of inputs x is estimated by $\widehat{\lambda}_{\text{VRS}}(x, y \mid \mathcal{S}_n)y$.

These ideas extend naturally to the input orientation as well as to the directional case. In both cases, the resulting estimators based on either $\widehat{\Psi}_{\text{VRS}}$ or $\widehat{\Psi}_{\text{CRS}}$ can be written as linear programs. In the hyperbolic case, $\widehat{\gamma}_{\text{CRS}}(x, y \mid \mathcal{S}_n)$ can be computed as the square root of $\widehat{\lambda}_{\text{CRS}}(x, y \mid \mathcal{S}_n)$ due to (2.8) in situations where Assumption 2.9(i) is maintained. Otherwise, Wilson (2011) provides a numerical method for computing the hyperbolic estimator $\widehat{\gamma}_{\text{VRS}}(x, y \mid \mathcal{S}_n)$ based on $\widehat{\Psi}_{\text{VRS}}$.

Both the FDH and DEA estimators are biased by construction since $\widehat{\Psi}_{\text{FDH}} \subseteq \widehat{\Psi}_{\text{VRS}} \subseteq \Psi$. Moreover, $\widehat{\Psi}_{\text{VRS}} \subseteq \widehat{\Psi}_{\text{CRS}}$. Under Assumption 2.9(i), $\widehat{\Psi}_{\text{CRS}} \subseteq \Psi$; otherwise, $\widehat{\Psi}_{\text{CRS}}$ will not be a statistically consistent estimator of Ψ. Of course, if Ψ is not convex, then $\widehat{\Psi}_{\text{VRS}}$ will also be inconsistent. These relations further imply that $\widehat{\lambda}_{\text{FDH}}(x, y \mid \mathcal{S}_n) \leq \widehat{\lambda}_{\text{VRS}}(x, y \mid \mathcal{S}_n) \leq$

$\lambda(x, y \mid \Psi)$ and $\widehat{\lambda}_{\text{VRS}}(x, y \mid \mathcal{S}_n) \leq \widehat{\lambda}_{\text{CRS}}(x, y \mid \mathcal{S}_n)$. Similar relations hold for estimators of the input, hyperbolic and directional efficiency measures.

Kneip et al. (1998) derive the rate of convergence of the input-oriented DEA estimator, while Kneip et al. (2008) derive its limiting distribution. These results extend to the output orientation after straightforward (though perhaps tedious) changes in notation to establish that for $(x, y) \in \Psi$, and conditions satisfied by the assumptions listed in Section 2,

$$n^{2/(p+q+1)} \left(\widehat{\lambda}_{\text{VRS}}(x, y \mid \mathcal{S}_n) - \lambda(x, y \mid \Psi) \right) \xrightarrow{\mathcal{L}} Q_{xy,\text{VRS}}(\cdot) \tag{3.11}$$

as $n \to \infty$ where $Q_{xy,\text{VRS}}(\cdot)$ is a regular, non-degenerate distribution with parameters depending on the characteristics of the DGP and on (x, y). Wilson (2011) establishes similar results for the hyperbolic orientation, and Simar et al. (2012) extend these results to the directional case. The convergence rate remains the same in both cases. Under similar assumptions but with the addition of CRS in Assumption 2.9, Park et al. (2010) establish

$$n^{2/(p+q)} \left(\widehat{\lambda}_{\text{CRS}}(x, y \mid \mathcal{S}_n) - \lambda(x, y \mid \Psi) \right) \xrightarrow{\mathcal{L}} Q_{xy,\text{CRS}}(\cdot) \tag{3.12}$$

for $(x, y) \in \Psi$ as $n \to \infty$, where $Q_{xy,\text{CRS}}(\cdot)$ is another regular, non-degenerate distribution with parameters depending on the characteristics of the DGP and on (x, y) but different from $Q_{xy,1}(\cdot)$. Interestingly, results from Kneip et al. (2016) for the input-oriented case can be extended to the output orientation (again, after straightforward but tedious changes in notation) to establish that under Assumption 2.9 and appropriate regularity conditions (see Kneip et al., 2016 for details),

$$\widehat{\lambda}_{\text{VRS}}(x, y \mid \mathcal{S}_n) - \lambda(x, y \mid \Psi) = O_P(n^{-2/(p+q)}). \tag{3.13}$$

Thus, under CRS, the VRS-DEA estimator attains the faster rate of the CRS-DEA estimator. The FDH estimator, however, keeps the same convergence rate $n^{1/(p+q)}$ regardless of returns to scale or whether Ψ is convex.

Kneip et al. (2015b) provide results on the moments of both FDH and DEA estimators as discussed below in Section 5. Limiting distributions of the FDH efficiency estimators have a Weibull form, but with parameters that are difficult to estimate. that are difficult to estimate. The limiting distributions of the DEA estimators do not have a closed form. Hence in either case, inference on individual efficiency scores requires bootstrap techniques. In the

DEA case, Kneip et al. (2008) provide theoretical results for both a smoothed bootstrap and for subsampling, while Kneip et al. (2011) and Simar and Wilson (2011b) provide details and methods for practical implementation. Subsampling can also be used for inference in the FDH case; see Jeong and Simar (2006) and Simar and Wilson (2011b).

3.3 Order-m Estimators

In Sections 3.1 and 3.2, the plug-in approach was used to define estimators of the output efficiency measure defined in (2.5) based on the full frontier. A similar approach is used here to define an estimator of $\lambda_m(x, y \mid H_{XY})$ defined in (2.18). The empirical analog of the distribution function $H_{XY}(x, y)$ defined in (2.11) is

$$\widehat{H}_{XY,n}(x, y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x, Y_i \geq y), \tag{3.14}$$

where $\mathbb{1}(\cdot)$ is the indicator function (i.e., $\mathbb{1}(A) = 1$ if A; otherwise, $\mathbb{1}(A) = 0$). Note that at any point $(x, y) \in \mathbb{R}^{p+q}$, $\widehat{H}_{XY,n}(x, y)$ gives the proportion of sample observations in $\mathcal{S}_n$ with values $X_i \leq x$ and $Y_i \geq y$; in other words, $\widehat{H}_{XY,n}(x, y)$ gives the proportion of points in the sample $\mathcal{S}_n$ that (weakly) dominate (x, y). In addition,

$$\widehat{S}_{Y|X,n}(y \mid x) = \frac{\widehat{H}_{XY,n}(x, y)}{\widehat{H}_{XY,n}(x, 0)} \tag{3.15}$$

is the empirical analog of the conditional survivor function $S_{Y|X}(y \mid x)$ defined in (2.13).

Substitution of (3.15) into the second line of (2.18) yields a nonparametric estimator of $\lambda_m(x, y \mid H_{XY})$, namely

$$\lambda_{m,n}(x, y \mid \widehat{H}_{XY,n}) = \int_0^\infty \left[1 - (1 - \widehat{S}_{Y,n}(uy \mid x))^m\right] du. \tag{3.16}$$

The integral in (3.16) is univariate and can be evaluated using numerical integration methods such as Gaussian quadrature. Alternatively, the estimator can be computed using Monte Carlo methods as described by Cazals et al. (2002). For a firm operating at $(x, y) \in \mathcal{P}$, an estimate of its expected maximum output level of order-m is given by

$$\widehat{y}_m^{\partial}(x) = \lambda(x, y \mid \widehat{H}_{XY,n})y, \tag{3.17}$$

analogous to (2.20).

Cazals et al. (2002) show that for finite m, $\lambda_m(x, y \mid \widehat{H}_{XY,n})$ is asymptotically normal with root-n convergence rate; i.e.,

$$\sqrt{n}\left(\lambda_m(x, y \mid \widehat{H}_{XY,n}) - \lambda_m(x, y)\right) \xrightarrow{\mathcal{L}} N(0, \sigma_m^2(x, y)) \tag{3.18}$$

as $n \to \infty$. An expression for the variance term σ_m^2 is given by Cazals et al. (2002). Although the convergence rate does not depend on the dimensionality (i.e., on $p + q$), the variance increases with $p + q$ when the sample size n is held fixed. It is not hard to find examples where most if not all observations in a sample lie above the order-m frontier unless m is very large.

Cazals et al. (2002) discuss the input oriented analog of the estimator in (3.16). Wilson (2011) extends these ideas to the hyperbolic case, and Simar and Vanhems (2012) extend the ideas to directional distances. In each case, the estimators achieve strong consistency with root-n rate of convergence and asymptotic normality when m is finite. In addition, as $m \to \infty$, $\lambda_m(x, y) \to \lambda(x, y \mid \Psi)$ and $\lambda_{m,n}(x, y \mid \widehat{H}_{XY,n}) \to \widehat{\lambda}_{FDH}(x, y\mathcal{S}_n)$. If $m = m(n) \to \infty$ at rate $n \log nF_X(x)$ as $n \to \infty$,

$$n^{\frac{1}{p+q}}\left[\lambda_{m,n}(x, y \mid \widehat{H}_{XY,n}) - \lambda(x, y \mid \Psi)\right] \xrightarrow{\mathcal{L}} \text{Weibull}(\mu_{x,y}, p + q) \text{ as } n \to \infty. \tag{3.19}$$

As $m = m(n) \to \infty$, the estimator $\lambda_{m,n}(x, y \mid \widehat{H}_{XY,n})$ shares the same properties as the FDH estimator. But, in finite samples, it will be more robust to outliers and extreme values since it will not envelop all the observations in the sample.

3.4 Order-α Estimators

Here also the plug-in principle can be used to obtain nonparametric order-α efficiency estimators. A nonparametric estimator of the output α-quantile efficiency score defined in (2.22) is obtained by replacing the conditional survival function in (2.22) with its empirical analog introduced above in Section 3.3, yielding

$$\lambda_\alpha(x, y \mid \widehat{H}_{XY,n}) = \sup\{\lambda \mid \widehat{S}_{Y|X,n}(\lambda y \mid x) > 1 - \alpha\}. \tag{3.20}$$

The order-α efficiency estimator in (3.20) is easy to compute using the algorithm given by Daouia and Simar (2007). Define

$$\mathcal{Y}_i = \min_{k=1,\,\ldots,\,q} \frac{Y_i^k}{y^k}, \tag{3.21}$$

$i = 1, \ldots, n$, and let $n_x = n\widehat{F}_X(x) > 0$ where $\widehat{F}_X(x)$ is the p-variate empirical distribution function of the input quantities, i.e., the empirical analog of $F_X(x)$ defined by (2.13). For $j = 1, \ldots, n_x$, let $\mathcal{Y}^x_{(j)}$ denote the jth order statistic of the observations $\mathcal{Y}_i$ such that $X_i \leq x$, so that $\mathcal{Y}^x_{(1)} \leq \mathcal{Y}^x_{(2)} \leq \ldots \leq \mathcal{Y}^x_{(n_x)}$. Then

$$\widehat{\lambda}_{\alpha,n} = \begin{cases} \mathcal{Y}^x_{(\alpha n_x)} & \text{if } \alpha n_x \in \mathbb{N}_{++}; \\ \mathcal{Y}^x_{(\lfloor \alpha n_x \rfloor + 1)} & \text{otherwise,} \end{cases} \tag{3.22}$$

where $\lfloor \alpha n_x \rfloor$ denotes the integer part of αn_x and $\mathbb{N}_{++}$ is the set of strictly positive integers.

The input-oriented estimator is obtained similarly; see Daouia and Simar (2007) for details. The ideas are extended to the hyperbolic case by Wheelock and Wilson (2008), who also present a numerical procedure for computing the estimator that is faster than the method based on sorting by a factor of about 70. The directional case is treated by Simar and Vanhems (2012).

Regardless of the direction that is chosen, provided $\alpha < 1$, the order-α estimators converge completely; e.g., in the output orientation,

$$\lambda_{\alpha,n}(x, y \mid H_{XY,n}) \xrightarrow{c} \lambda_\alpha(x, y), \tag{3.23}$$

where $\xrightarrow{c}$ denotes complete convergence. Complete convergence, introduced by Hsu and Robbins (1947), implies and is a stronger form of convergence than almost-sure convergence. A sequence of random variables $\{\zeta_n\}_{n=1}^\infty$ converges completely to a random variable ζ, denoted by $\zeta_n \xrightarrow{c} \zeta$, if $\lim_{n\to\infty} \sum_{j=1}^n \Pr\left(|\zeta_j - \zeta| \geq \epsilon\right) < \infty \; \forall \; \epsilon > 0$. The complete convergence in (3.23) implies

$$\lim_{n\to\infty} \sum_{j=1}^{n} \Pr\left(|\lambda_{\alpha,j}(x, y \mid \widehat{H}_{XY,n}) - \lambda_\alpha(x, y)| \geq \epsilon\right) < \infty \tag{3.24}$$

for all $\epsilon > 0$. In addition, for $\alpha < 1$, the order-α estimator is asymptotically normally distributed, with root-n convergence rate:

$$\sqrt{n}(\lambda_{\alpha,n}(x, y \mid \widehat{H}_{XY,n}) - \lambda_\alpha(x, y)) \xrightarrow{\mathcal{L}} N(0, \sigma^2_\alpha(x, y)), \quad \text{as } n \to \infty. \tag{3.25}$$

An expression is for the variance term $\sigma^2_\alpha(x, y)$ is given by Daouia and Simar (2007). Note, however, that similar to the order-m estimator, as $\alpha \to 1$ the order-α estimator $\lambda_{\alpha,n}(x, y \mid \widehat{H}_{XY,n})$ shares the properties of the FDH estimator.

3.5 Making Inference About Efficiency of a Firm

The root-n rate of convergence of the order-m and order-α estimators is unusual among nonparametric estimators. Asymptotic normality of the partial efficiency estimators facilitates inference and allows estimation of confidence intervals using an asymptotic normal approximation. In the case of the order-m estimator, Cazals et al. (2002) describe a simple plug-in method for estimating the variance parameter $\sigma_m^2(x, y)$ that appears in (3.18). A similar approach can be used to estimate the variance term $\sigma_\alpha^2(x, y)$ that appears in (3.25). In either case, due to the asymptotic normality of the order-m and order-α estimators, the variance terms can also be estimated using a standard, naive bootstrap where observations are resampled uniformly, with replacement to create bootstrap samples of size n.

When either FDH or DEA estimators are used, making inference is more problematic. Although the limiting distributions of the FDH and DEA estimators have been derived, they are complicated. The limiting distribution of the FDH estimator is a Weibull, but its parameter depends on unknown model features (e.g., curvature of the true, unobserved frontier) that are difficult to estimate. In the DEA case, the limiting distributions do not have closed-form expressions. The only practical approach is to make inference using bootstrap methods. However, it is well-known (e.g., see Bickel and Freedman, 1981) that an ordinary, naive bootstrap does not provide valid inference when estimating a support boundary (or distance to the support boundary). Simar and Wilson, 2011a provide a pedagogical explanation of the problem. Unfortunately, this fact is not recognized in some of the frontier literature as indicated by the discussion in Simar and Wilson (1999a, 1999b).

Simar and Wilson (1998) propose a smooth bootstrap to replace the inconsistent naive bootstrap when FDH or DEA estimators are used. Simar and Wilson implement the smoothed bootstrap in a simple model under the assumption that the distribution of the inefficiencies along the chosen direction (input rays or output rays) is homogeneous in the input-output space. Hence the smoothing operates only on the estimation of the univariate density of the efficiencies, making the problem much easier to handle. Simar and Wilson (2000) extend this idea to a more general heterogeneous case where the distribution of efficiency is allowed to vary over Ψ. This requires more complication than the original procedure, and involves the estimation of a smoothed density of (X, Y) with unknown support in a $(p + q)$-dimensional space. No theoretical justification was given for either approach, but

results from intensive Monte-Carlo experiments described in both papers suggest that these bootstrap procedures give reasonable approximations for correcting the bias of the efficiency estimates and for building individual confidence intervals for the efficiency of any fixed point (x, y).

Kneip et al. (2008) describe two bootstrap techniques and prove that both provide consistent, valid inference. The first is a double-smooth bootstrap where in addition to smoothing the empirical distribution of the data, the support of Ψ is estimated by a smoothed version of the VRS-DEA estimator. The second is a subsampling bootstrap.

The double-smooth bootstrap developed by Kneip et al. (2008) involves numerical difficulties making it difficult to implement and computationally demanding. Kneip et al. (2011) provide a simplified, consistent, and computationally-efficient version of the double-smooth bootstrap. The idea is rather simple. It is well-known that the naive bootstrap does not work, but it does not work only because of points in a neighborhood of the boundary. The idea behind the simplified Kneip et al. (2011) method is to draw naively among observations which are "far" from the frontier and draw the remaining points from a uniform distribution with support "near" the frontier. This neighborhood of the frontier is tuned by a smoothing parameter that can be selected by a simple "rule of thumb." For obtaining consistency, the VRS-DEA frontier estimate must be smoothed, and here a second bandwidth parameter is selected by cross-validation methods.

The subsampling approach is much easier to implement, since a bootstrap sample $\mathcal{S}_{\widetilde{n}}^*$, where $\widetilde{n} = n^\gamma$ for some $\gamma \in (0, 1)$, is obtained by drawing (either with or without replacement) $\widetilde{n}$ pairs (X_i, Y_i) from the original sample $\mathcal{S}_n$. Kneip et al. (2008) prove consistency of this subsampling bootstrap, but do not provide suggestions for how a value for $\widetilde{n}$ might be selected in practice. Their simulation results indicate that performance of the subsampling bootstrap in terms of achieved coverages of estimated confidence intervals is quite sensitive to the choice of $\widetilde{n}$. Jeong and Simar (2006) prove that subsampling provides a consistent approximation of the sampling distribution of the FDH efficiency estimator, but here again, no practical advice is offered on how to select an appropriate subsample size.

Using results from Politis et al. (2001) and Bickel and Sakov (2008), Simar and Wilson (2011a) provide a data-driven algorithm for selecting an appropriate value of the subsample size $\widetilde{n}$, for both the FDH and DEA cases. The idea is to compute the object of interest (e.g.,

bounds of a confidence interval, or bias estimate) for various values of $\widetilde{n}$ on some selected grid. Then the value of $\widetilde{n}$ where the results show the smallest volatility is selected. This volatility can be computed for each value of $\widetilde{n}$ in the grid by computing, for example, the standard deviation between the 3 or 5 values found for the adjacent values of $\widetilde{n}$. Simar and Wilson (2011a) investigate the performance of their method (in terms of achieved coverages of individual confidence intervals for efficiency scores) by intensive Monte-Carlo experiments, for both FDH and DEA estimators. The results indicate that the method works well for moderate sample sizes similar to those faced in practice, providing reasonable approximations of the sampling distribution of the estimators. The method has been implemented in the FEAR software package described by Wilson (2008).

4 Environmental Factors

4.1 A Statistical Model with Environmental Variables

The statistical model presented in Section 2 include only input quantities and output quantities represented by random variables $X \in \mathbb{R}^p_+$ and $Y \in \mathbb{R}^q_+$, respectively. But in some situations, the production process may be affected by influences beyond inputs and outputs. For example, firms may face different regulatory environments. In the US hospital industry, hospital type or ownership (i.e., for-profit, non-profit, state and local government, or federal government) may influence incentives, objectives, or other features that might affect production. Similarly, in banking and insurance, private, public, or mutual ownership might influence production. In agricultural studies, rainfall might be regarded as something other than an input, exogenous to farmers' choices. Below, we refer to such factors—that are neither inputs nor outputs—as *environmental factors,* and formalize a statistical model of the production process along the lines of the probability framework of Cazals et al. (2002) and Daraio et al. (2018).

Let $Z \in \mathbb{R}^r$ denote a vector of variables describing environmental factors faced by producers, with random variables (X, Y, Z) defined on an appropriate probability space. Let $f_{XYZ}(x, y, z)$ denote the joint density of (X, Y, Z) which has support $\mathcal{P} \subset \mathbb{R}^p_+ \times \mathbb{R}^q_+ \times \mathbb{R}^r$. This joint density can always be decomposed as

$$f_{XYZ}(x, y, z) = f_{XY|Z}(x, y \mid z) f_Z(z). \tag{4.1}$$

Let

$$\Psi^z = \{(X, Y) \mid X \text{ can produce } Y \text{ when } Z = z\} \tag{4.2}$$

be the set of feasible combinations of inputs and outputs for a firm facing environmental conditions $Z = z$. Then Ψ^z is the conditional support of $f_{XY|Z}(x, y \mid z)$, i.e., the support of (x, y) given $Z = z$. Let $\mathcal{Z}$ denote the support of $f_Z(z)$.

There are several ways in which the environmental variables in Z might affect the production process. First, the environmental variables might affect only the inefficiency process through the density $f_{XY|Z}(x, y \mid z)$, thereby affecting the probability for a firm to be within some neighborhood of the frontier. Second, the environmental variables might operate only through Ψ^z, thereby determining the support of (X, Y) but not the level of inefficiency. In other words, the environmental variables might affect production possibilities, but not the level of inefficiency. A third possibility is that the variables in Z affect both the level of inefficiency as well as the support of (X, Y).

Let

$$\Psi^+ = \bigcup_{z \in \mathcal{Z}} \Psi^z. \tag{4.3}$$

By construction, $\Psi^z \subseteq \Psi^+ \subset \mathbb{R}_+^{p+q} \; \forall \; z \in \mathcal{Z}$. Then *one and only one* of the following conditions must be satisfied.

Assumption 4.1. *(Separability Condition):* $\Psi^z = \Psi^+$ *for all* $z \in \mathcal{Z}$.

Assumption 4.2. *(Non Separability Assumption):* $\Psi^z \neq \Psi^+$ *for some* $z \in \mathcal{Z}$.

Assumptions 4.1 and 4.2 are mutually exclusive and collectively exhaustive. Assumption 4.1 is referred to as the "separability" condition by Simar and Wilson (2007).

Under Assumption 4.1 it is clear that the joint support $\mathcal{P}$ of (X, Y, Z) can be factorized as

$$\mathcal{P} = \Psi^+ \times \mathcal{Z}, \tag{4.4}$$

and Ψ^+ is equivalent to Ψ defined by (2.1).[10] However, $\Psi^+ = \Psi$ *if and only if* Assumption 4.1 holds. Alternatively, under Assumption 4.2, Ψ is not well-defined since in this case

[10] Assumption 4.1 is referred to as the "separability condition" by Simar and Wilson (2007) because of (4.4), i.e., since the support of (X, Y, Z) can be written as the Cartesian product of the production set $\Psi^+ = \Psi$ and the support of Z when Assumption 4.1 holds.

whether X can produce Y depends necessarily depends upon Z. Under Assumption 4.2, Ψ^+ remains well-defined but contains pairs (X, Y) that are not feasible. Moreover, under Assumption 4.2, Ψ^+ does not describe any useful feature of the model and has no useful economic interpretation.

Under Assumption 4.1, one can perform a two-stage analysis where the environmental variables in Z are ignored in the first stage and efficiency is estimated by one of the *unconditional* estimators $\widehat{\lambda}_{\text{FDH}}(x, y \mid \mathcal{S}_n)$, $\widehat{\lambda}_{\text{VRS}}(x, y \mid \mathcal{S}_n)$, or $\widehat{\lambda}_{\text{CRS}}(x, y \mid \mathcal{S}_n)$. Then the set of resulting efficiency estimates can be regressed in a second-stage on the environmental variables Z using a truncated regression model as explained by Simar and Wilson (2007). Note, however, that conventional inference (e.g., using estimates of variance obtained by inverting the negative Hessian of the log-likelihood for the truncated regression) is invalid in the second-stage regression. Simar and Wilson (2007) propose a bootstrap method for making inference in the second stage; see Simar and Wilson (2011b) and Kneip et al. (2015b) for additional discussion regarding inference-making in the second stage regression.

Alternatively, if efficiency is estimated in a first stage using one of the *unconditional* efficiency estimators and then the resulting efficiency estimates are regressed on elements of Z, then neither the first-stage estimates nor the second-stage estimates estimate any useful or meaningful model feature if Assumption 4.2 holds instead of Assumption 4.1. Since Assumption 4.1 is a rather restrictive assumption, care should be taken. Simar and Wilson (2007) note that Assumption 4.1 should be tested against the alternative hypothesis given by Assumption 4.2, but such a test has only recently been provided by Daraio et al. (2018). The test is discussed below in Section 6.4.

Since Ψ is not well-defined under Assumption 4.2, the efficiency measures defined in (2.3), (2.5), (2.7) and (2.9) are also not well-defined. Whenever Assumption 4.2 holds, notions of *conditional* efficiency are needed. In the output orientation, the conditional measure

$$\lambda(x, y \mid z) = \sup\{\lambda > 0 \mid (x, \lambda y) \in \Psi^z\} \tag{4.5}$$

introduced by Cazals et al. (2002) and Daraio and Simar (2005) gives a measure of distance to the appropriate, relevant boundary (i.e., the boundary of Ψ^z that is attainable by firms operating under conditions described by z).

As in Section 2.3, the conditional efficiency measure in (4.5) can be described in probabilistic terms. The conditional density $f_{XY|Z}(x, y \mid z)$ in (4.1) implies a corresponding

distribution function

$$H_{XY|Z}(x, y \mid z) = \Pr(X \leq x,\ Y \geq y \mid Z = z), \tag{4.6}$$

giving the probability of finding a firm dominating the production unit operating at the level (x, y) and facing environmental conditions z. Then analogous to the reasoning in Section 2.3, the conditional efficiency score can be written as

$$\lambda(x, y \mid z) = \sup\{\lambda > 0 \mid H_{XY|Z}(x, \lambda y \mid z) > 0\}. \tag{4.7}$$

Let h denote a vector of bandwidths of length r corresponding to elements of Z and z. In order to define a statistical model that incorporates environmental variables, consider an h-neighborhood of $z \in \mathcal{Z}$ such that $|Z - z| \leq h$ and define the *conditional* attainable set given by

$$\begin{aligned}\Psi^{z,h} &= \{(X, Y) \mid X \text{ can produce } Y, \text{ when } |Z - z| \leq h\} \\ &= \left\{(x, y) \in \mathbb{R}_+^{p+q} \mid H^h_{XY|Z}(x, y \mid z) > 0\right\} \\ &= \left\{(x, y) \in \mathbb{R}_+^{p+q} \mid f^h_{XY|Z}(\cdot, \cdot \mid z) > 0\right\}\end{aligned} \tag{4.8}$$

where

$$H^h_{XY|Z}(x, y \mid z) = \Pr(X \leq x, Y \geq y \mid z - h \leq Z \leq z + h) \tag{4.9}$$

gives the probability of finding a firm dominating the production unit operating at the level (x, y) and facing environmental conditions Z in an h-neighborhood of z. The corresponding conditional density of (X, Y) given $|Z - z| \leq h$, denoted by $f^h_{XY|Z}(\cdot, \cdot \mid z)$, is implicitly defined by

$$H^h_{XY|Z}(x, y \mid z) = \int_{-\infty}^{x} \int_{y}^{\infty} f^h_{XY|Z}(u, v \mid Z \in [z - h, z + h])\, dv\, du. \tag{4.10}$$

Clearly, $\Psi^{z,h} = \bigcup_{|\tilde{z} - z| \leq h} \Psi^{\tilde{z}}$. Following Jeong et al. (2010), define for a given h

$$\begin{aligned}\lambda^h(x, y \mid z) &= \sup\{\lambda > 0 \mid (x, \lambda y) \in \Psi^{z,h}\} \\ &= \sup\left\{\lambda > 0 \mid H^h_{XY|Z}(x, \lambda y \mid z) > 0\right\}.\end{aligned} \tag{4.11}$$

The idea behind estimating efficiency *conditionally* on $Z = z$ is to let the bandwidth h tend to 0 as $n \to \infty$. The idea is motivated by smoothing methods similar to those used

in nonparametric regression problems, but adapted to frontier estimation. Some additional, new assumptions are needed to complete the statistical model.

The next three assumptions are conditional analogs of Assumptions 2.1–2.3 appearing above in Section 2.1.

Assumption 4.3. *For all $z \in \mathcal{Z}$, Ψ^z and $\Psi^{z,h}$ are closed.*

Assumption 4.4. *For all $z \in \mathcal{Z}$, both inputs and outputs are strongly disposable; i.e., for any $z \in \mathcal{Z}$, $\widetilde{x} \geq x$ and $0 \leq \widetilde{y} \leq y$, if $(x, y) \in \Psi^z$ then $(\widetilde{x}, y) \in \Psi^z$ and $(x, \widetilde{y}) \in \Psi^z$. Similarly, if $(x, y) \in \Psi^{z,h}$ then $(\widetilde{x}, y) \in \Psi^{z,h}$ and $(x, \widetilde{y}) \in \Psi^{z,h}$.*

Assumption 4.5. *For all $z \in \mathcal{Z}$, if $x = 0$ and $y \geq 0$, $y \neq 0$ then (i) $(x, y) \notin \Psi^z$ and (ii) $(x, y) \notin \Psi^{z,h}$.*

Assumption 4.4 corresponds to Assumption 1F in Jeong et al. (2010), and amounts to a regularity condition on the conditional attainable sets justifying the use of the localized versions of the FDH and DEA estimators. The assumption imposes weak monotonicity on the frontier in the space of inputs and outputs for a given $z \in \mathcal{Z}$, and is standard in micro-economic theory of the firm. Assumption 4.5 is the conditional analog of Assumption 2.3 and rules out free lunches.

The next assumption concerns the regularity of the density of Z and of the conditional density of (X, Y) given $Z = z$, as a function of z in particular near the efficient boundary of Ψ^z (see Assumptions 3 and 5 in Jeong et al., 2010).

Assumption 4.6. *Z has a continuous density $f_Z(\cdot)$ such that for all $z \in \mathcal{Z}$ $f_Z(z)$ is bounded away from zero. Moreover the conditional density $f_{XY|Z}(\cdot, \cdot \mid z)$ is continuous in z and is strictly positive in a neighborhood of the frontier of Ψ^z.*

Assumption 4.7. *For all (x, y) in the support of (X, Y), $\lambda^h(x, y \mid z) - \lambda(x, y \mid z) = O(h)$ as $h \to 0$.*

Assumption 4.7 amounts to an assumption of continuity of $\lambda(\cdot, \cdot \mid z)$ as a function of z, and is analogous to Assumption 2 of Jeong et al. (2010).

The remaining assumptions impose regularity conditions on the data-generating process. The first assumption appears as Assumption 4 in Jeong et al. (2010).

Assumption 4.8. *(i) The sample observations* $\mathcal{X}_n = \{(X_i, Y_i, Z_i)\}_{i=1}^n$ *are realizations of iid random variables* (X, Y, Z) *with joint density* $f_{XYZ}(\cdot,\cdot,\cdot)$*; and (ii) the joint density* $f_{XYZ}(\cdot,\cdot,\cdot)$ *of* (X, Y, Z) *is continuous on its support.*

The next assumptions are needed to establish results for the moments of the conditional FDH and DEA estimators described in Section 4.2. The assumptions here are conditional analogs of Assumptions 3.1–3.4 and 3.6 (respectively) in Kneip et al. (2015b). Assumption 4.9, part (iii) and Assumption 4.10, part(iii) appear as Assumption 5 in Jeong et al. (2010).

Assumption 4.9. *For all* $z \in \mathcal{Z}$*, (i) the conditional density* $f_{XY|Z}(\cdot,\cdot \mid z)$ *of* $(X, Y) \mid Z = z$ *exists and has support* $\mathcal{D}^z \subset \Psi^z$*; (ii)* $f_{XY|Z}(\cdot,\cdot \mid z)$ *is continuously differentiable on* $\mathcal{D}^z$*; and (iii)* $f^h_{XY|Z}(\cdot,\cdot \mid z)$ *converges to* $f_{XY|Z}(\cdot,\cdot \mid z)$ *as* $h \to 0$*.*

Assumption 4.10. *(i)* $\mathcal{D}^{z*} := \{(x, \lambda(x, y \mid z)y) \mid (x, y) \in \mathcal{D}^z\} \subset \mathcal{D}^z$*; (ii)* $\mathcal{D}^{z*}$ *is compact; and (iii)* $f_{XY|Z}(x, \lambda(x, y \mid z)y \mid z) > 0$ *for all* $(x, y) \in \mathcal{D}^z$*.*

Assumption 4.11. *For any* $z \in \mathcal{Z}$*,* $\lambda(x, y \mid z)$ *is three times continuously differentiable with respect to* x *and* y *on* $\mathcal{D}^z$*.*

Assumption 4.12. *For all* $z \in \mathcal{Z}$*, (i)* $\lambda(x, y \mid z)$ *is twice continuously differentiable on* $\mathcal{D}^z$*; and (ii) all the first-order partial derivatives of* $\lambda(x, y \mid z)$ *with respect to* x *and* y *are nonzero at any point* $(x, y) \in \mathcal{D}^z$*.*

Assumption 4.13. *For any* $z \in \mathcal{Z}$*,* $\mathcal{D}^z$ *is almost strictly convex; i.e., for any* (x, y)*,* $(\widetilde{x}, \widetilde{y}) \in \mathcal{D}^z$ *with* $\left(\frac{x}{||x||}, y\right) \neq \left(\frac{\widetilde{x}}{||\widetilde{x}||}, \widetilde{y}\right)$*, the set* $\{(x^*, y^*) \mid (x^*, y^*) = (x, y) + \alpha((\widetilde{x}, \widetilde{y}))$ *for some* $\alpha \in (0, 1)\}$ *is a subset of the interior of* $\mathcal{D}^z$*.*

Assumption 4.14. *For any* $z \in \mathcal{Z}$*, (i) for any* $(x, y) \in \Psi^z$ *and any* $a \in [0, \infty)$*,* $(ax, ay) \in \Psi^z$*; (ii) the support* $\mathcal{D}^z \subset \Psi^z$ *of* $f_{XY|Z}$ *is such that for any* (x, y)*,* $(\widetilde{x}, \widetilde{y}) \in \mathcal{D}^z$ *with* $(\frac{x}{||x||}, \frac{y}{||y||}) \neq (\frac{\widetilde{x}}{||\widetilde{x}||}, \frac{\widetilde{y}}{||\widetilde{y}||})$*, the set* $\{(x^*, y^*) \mid (x^*, y^*) = (x, y) + \alpha((\widetilde{x}, \widetilde{y}) - (x, y))$ *for some* $0 < \alpha < 1\}$ *is a subset of the* interior *of* $\mathcal{D}^z$*; and (iii)* $(x, y) \notin \mathcal{D}^z$ *for any* $(x, y) \in \mathbb{R}^p_+ \times \mathbb{R}^q$ *with* $y^1 = 0$*, where* y^1 *denotes the first element of the vector* y*.*

When the conditional FDH estimator is used, Assumption 4.12 is needed; when the conditional DEA estimator is used, this is replaced by the stronger Assumption 4.11.

Note that Assumptions 4.3–4.5 and 4.8–4.14 for the model with environmental variables are analogs of Assumptions 2.1–2.3 and 2.4–2.9 for the the model without environmental variables. The set of Assumptions 4.8–4.14 are stronger than set of assumptions required by Jeong et al. (2010) to prove consistency and to derive the limiting distribution for conditional efficiency estimators. The stronger assumptions given here are needed by Daraio et al. (2018) to obtain results on moments of the conditional efficiency estimators as well as the central limit theorem (CLT) results discussed below in Section 5.2. Daraio et al. (2018) do not consider the conditional version of the CRS-DEA estimator, and hence do not use Assumption 4.14 that appears above. However, the results obtained by Daraio et al. (2018) for the conditional version of the VRS-DEA estimator that are outlined below in Section 4.2 can be extended to the CRS case while replacing Assumption 4.13 with Assumption 4.14. In addition, note that under the separability condition in Assumption 4.1, the assumptions here reduce to the corresponding assumptions in Kneip et al. (2015b) due to the discussion in Section 2.

4.2 Non-parametric Conditional Efficiency Estimators

As in previous cases, the plug-in approach is useful for defining estimators of the conditional efficiency score given in (4.5) and (4.7).

For the conditional efficiency score $\lambda(x, y \mid z)$, a smoothed estimator of $H_{XY|Z}(x, y \mid z)$ is needed to plug into (4.7), which requires a the vector h of bandwidths for Z. The conditional distribution function $H_{XY|Z}(x, y \mid z)$ can be replaced by the estimator

$$\widehat{H}_{XY|Z}(x, y \mid z) = \frac{\sum_{i=1}^{n} I(X_i \leq x, Y_i \geq y) K_h(Z_i - z)}{\sum_{i=1}^{n} K_h(Z_i - z)}, \tag{4.12}$$

where $K_h(\cdot) = (h_1 \ \ldots \ h_r)^{-1} K\left((Z_i - z)/h\right)$ and the division between vectors is understood to be component-wise. As explained in the literature (e.g., see Daraio and Simar, 2007b), the kernel function $K(\cdot)$ must have bounded support (e.g., the Epanechnikov kernel). This provides the output-oriented, conditional FDH estimator

$$\widehat{\lambda}_{\text{FDH}}(x, y \mid z, \mathcal{X}_n) = \max_{i \in \mathcal{I}_1(z,h)} \left(\min_{j=1, \ldots, p} \left(\frac{Y_i^j}{y^j} \right) \right), \tag{4.13}$$

where

$$\mathcal{I}_1(z, h) = \{ i \mid z - h \leq Z_i \leq z + h \ \cap \ X_i \leq x \} \tag{4.14}$$

is the set of indices for observations with Z in an h-neighborhood of z and for which output levels X_i are weakly less than x.

Alternatively, where one is willing to assume that the conditional attainable sets are convex, Daraio and Simar (2007b) suggest the conditional VRS-DEA estimator of $\lambda(x, y \mid z)$ given by

$$\widehat{\lambda}_{\text{VRS}}(x, y \mid z, \mathcal{X}_n) = \max_{\lambda, \omega_1, \ldots, \omega_n} \Big\{ \lambda > 0 \mid \lambda y \leq \sum_{i \in \mathcal{I}_2(z,h)} \omega_i Y_i,\ x \geq \sum_{i \in \mathcal{I}_2(z,h)} \omega_i X_i, \\ \text{for some } \omega_i \geq 0 \ \text{ such that } \sum_{i \in \mathcal{I}_2(z,h)} \omega_i = 1 \Big\} \tag{4.15}$$

where

$$\mathcal{I}_2(z, h) = \{ i \mid z - h \leq Z_i \leq z + h \} \tag{4.16}$$

is the set of indices for observations with Z in an h-neighborhood of z. Note that the conditional estimators in (4.13) and (4.15) are just localized version of the unconditional FDH and VRS-DEA efficiency estimators given in (3.4) and (3.10), where the degree of localization is controlled by the bandwidths in h. The conditional version of CRS-DEA estimator $\widehat{\lambda}_{\text{CRS}}(x, y \mid \mathcal{S}_n)$ is obtained by dropping the constraint $\sum_{i \in \mathcal{I}_2(z,h)} \omega_i = 1$ in (4.15) and is denoted by $\widehat{\lambda}_{\text{CRS}}(x, y \mid z, \mathcal{X}_n)$. Bandwidths can be optimized by least-squares cross validation; see Daraio et al. (2018) for discussion of practical aspects.

Jeong et al. (2010) show that the conditional version of the FDH and VRS-DEA efficiency estimators share properties similar to their unconditional counterparts whenever the elements of Z are continuous. The sample size n is replaced by the effective sample size used to build the estimates, which is of order $nh_1 \ \ldots \ h_r$, which we denote as n_h. To simplify the notation, and without loss of generality, we hereafter assume that all of the bandwidths $h_j = h$ are the same, so that $n_h = nh^r$. For a fixed point (x, y) in the interior of Ψ^z, as $n \to \infty$,

$$n_h^{\kappa} \Big(\widehat{\lambda}(x, y \mid z, \mathcal{X}_n) - \lambda(x, y \mid z) \Big) \xrightarrow{\mathcal{L}} Q_{xy|z}(\cdot) \tag{4.17}$$

where again $Q_{xy|z}(\cdot)$ is a regular, non-degenerate limiting distribution analogous to the corresponding one for the unconditional case. The main argument in Jeong et al. (2010) relies on on the property that there are enough points in a neighborhood of z, which is obtained with the additional assumption that $f_Z(z)$ is bounded away from zero at z and that if the

bandwidth is going to zero, it should not go too fast (see Jeong et al., 2010, Proposition 1; if $h \to 0$, h should be of order $n^{-\eta}$ with $\eta < r^{-1}$).

The conditional efficiency scores have also their robust versions; see Cazals et al. (2002) and Daraio and Simar (2007b) for the order-m version, and Daouia and Simar (2007) for the order-α analog. Also, conditional measures have been extended to hyperbolic distances in Wheelock and Wilson (2008), and to hyperbolic distances by Simar and Vanhems (2012).

Bădin et al. (2012, 2014) suggest useful tools for analyzing the impact of Z on the production process, by exploiting the comparison between the conditional and unconditional measures. These tools (graphical and nonparametric regressions) allow one to disentangle the impact of Z on any potential shift of the frontier or potential shift of the inefficiency distributions. Daraio and Simar (2014) provide also a bootstrap test for testing the significance of environmental factors on the conditional efficiency scores. These tools have been used in macroeconomics to gauge the effect of foreign direct investment and time on "catching-up" by developing countries; see Mastromarco and Simar (2015).

Florens et al. (2014) propose an alternative approach for estimating conditional efficiency scores that avoids explicit estimation of a nonstandard conditional distribution (e.g., $F_{X|Y,Z}(x \mid y, z)$). The approach is less sensitive to the curse of the dimensionality described above. It is based on very flexible nonparametric location-scale regression models for prewhitening the inputs and the outputs to eliminate their dependence on Z. This allows one to define "pure" inputs and outputs, and hence a "pure" measure of efficiency. The method permits returning in a second stage to the original units and evaluating the conditional efficiency scores, but without explicitly estimating a conditional distribution function. The paper proposes also a bootstrap procedure for testing the validity of the location-scale hypothesis. The usefulness of the approach is illustrated using data on commercial banks to analyze the effects of banks' size and diversity of the services offered on the production process, and on the resulting efficiency distribution.

As a final remark, note that if Assumption 4.1 holds, so that the variables in Z have no effect on the frontier, then Assumptions 4.3–4.11 can be shown to be equivalent to the corresponding conditions in Assumptions 2.1–2.9. As a practical matter, whenever Assumption 4.1 holds, least squares cross-validation will result in bandwidths large enough so that the sets of indices $\mathcal{I}_1(z, h)$ and $\mathcal{I}_2(z, h)$ includes all integers $1,\ 2,\ \ldots,\ n$. In this case, the vari-

ables in Z do not affect the frontier and the conditional efficiency estimators are equivalent to the corresponding unconditional estimators.

5 Central Limit Theorems for Mean Efficiency

5.1 Mean Unconditional Efficiency

CLT results are among the most fundamental, important results in statistics and econometrics; see Spanos (1999) for detailed discussion of their historical development and their importance in inference-making. CLTs are needed for making inference about population means. In the frontier context, one might want to make inference about

$$\begin{aligned}\mu_\lambda &= E\left(\lambda(X,Y \mid \Psi)\right) \\ &= \int_{\mathcal{D}} \lambda(x,y \mid \Psi) f(x,y)\, dx\, dy. \end{aligned} \tag{5.1}$$

If $\lambda(X_i, Y_i \mid \Psi)$ were observed for each $(X_i, Y_i) \in \mathcal{S}_n$, then μ could be estimated by the sample mean

$$\overline{\lambda}_n = \sum_{i=1}^{n} \lambda(X_i, Y_i \mid \Psi). \tag{5.2}$$

Then under mild conditions, the Lindeberg-Feller CLT establishes the limiting distribution of $\overline{\lambda}_n$, i.e.,

$$\sqrt{n}\left(\overline{\lambda}_n - \mu_\lambda\right) \xrightarrow{\mathcal{L}} N\left(0, \sigma_\lambda^2\right) \tag{5.3}$$

provided

$$\begin{aligned}\sigma_\lambda^2 &= \mathrm{VAR}\left(\lambda(X,Y \mid \Psi)\right) \\ &= \int_{\mathcal{D}} \left(\lambda(x,y \mid \Psi) - \mu_\lambda\right)^2 f(x,y)\, dx\, dy \end{aligned} \tag{5.4}$$

is finite. If the $\lambda(X_i, Y_i \mid \Psi)$ were observed for $i = 1, \ldots, n$, one could use (5.3) to estimate confidence intervals for μ_λ in the usual way, relying on asymptotic approximation for finite samples. But of course the $\lambda(X_i, Y_i \mid \Psi)$ are *not* observed, and must be estimated. Kneip et al. (2015b) show that the bias of the FDH and DEA estimators makes inference about μ_λ problematic.[11]

[11] Kneip et al. (2015b) focus on the input orientation, but all of their results extend to the output orientation after straightforward (but perhaps tedious) changes in notation. The discussion here is in terms of the output orientation.

Kneip et al. (2015b) establish results for the moments of FDH and DEA estimators under appropriate assumptions given in Section 2[12] These results are summarized by writing

$$E\left(\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n) - \lambda(X_i, Y_i)\right) = Cn^{-\kappa} + R_{n,\kappa}, \tag{5.5}$$

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

$$E\left(\left(\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n) - \lambda(X_i, Y_i)\right)^2\right) = o\left(n^{-\kappa}\right), \tag{5.6}$$

and

$$\left|\text{COV}\left(\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n) - \lambda(X_i, Y_i),\ \widehat{\lambda}(X_j, Y_j \mid \mathcal{S}_n) - \lambda(X_j, Y_j)\right)\right| = o\left(n^{-1}\right) \tag{5.7}$$

for all $i, j \in \{1, \ldots, n\}$, $i \neq j$. Here, we suppress the labels "VRS," "CRS," or "FDH" on $\widehat{\lambda}$. The values of the constant C, the rate κ, and the remainder term $R_{n,\kappa}$ depend on which estimator is used. In particular,

- under VRS with the VRS-DEA estimator, $\kappa = \frac{2}{(p+q+1)}$ and $R_{n,\kappa} = O(n^{-3\kappa/2}(\log n)^{\alpha_1})$, where $\alpha_1 = (p+q+4)/(p+q+1)$;
- under CRS with either the VRS-DEA or CRS-DEA estimator, $\kappa = \frac{2}{(p+q)}$ and $R_{n,\kappa} = O(n^{-3\kappa/2}(\log n)^{\alpha_2})$ where $\alpha_2 = (p+q+3)/(p+q)$; and
- under only the free disposability assumption (but not necessarily CRS or convexity) with the FDH estimator, $\kappa = \frac{1}{(p+q)}$ and $R_{n,\kappa} = O(n^{-2\kappa}(\log n)^{\alpha_3})$, where $\alpha_3 = (p+q+2)/(p+q)$.

Note that in each case, $R_{n,\kappa} = o\left(n^{-\kappa}\right)$.

The result in (5.7) is somewhat surprising. It is well-known that FDH and DEA estimates are correlated, due to the fact that typically the estimated efficiency of several, perhaps many observations depends on a small number of observations lying on the estimated frontier; i.e., perturbing an observation lying on the estimated frontier is likely to affect estimated efficiency for other observations. The result in (5.7) however indicates that this effect is negligible.

[12] Throughout this section, assumptions common to VRS-DEA, CRS-DEA, and FDH estimators include Assumptions 2.1–2.3 and Assumptions 2.4 and 2.5. For the VRS-DEA estimator under VRS, "appropriate assumptions" include the common assumptions as well as Assumptions 2.6 and 2.8. For the CRS-DEA estimator, "appropriate assumptions" consist of the common assumptions and Assumption 2.9. For the FDH estimator, "appropriate assumptions" consist of the common assumptions and Assumption 2.7.

The $Cn^{-\kappa}$ term in (5.5) reflects the bias of the nonparametric efficiency estimators, and its interplay with the $o\left(n^{-\kappa}\right)$ expression in (5.6) creates problems for inference. Let

$$\widehat{\mu}_n = n^{-1} \sum_{i=1}^{n} \widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n). \tag{5.8}$$

Theorem 4.1 of Kneip et al. (2015b) establishes that $\widehat{\mu}_n$ is a consistent estimator of μ under the appropriate set of assumptions, but has bias of order $Cn^{-\kappa}$. The theorem also establishes that

$$\sqrt{n}\left(\widehat{\mu}_n - \mu_\lambda - Cn^{-\kappa} - R_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0, \sigma_\lambda^2). \tag{5.9}$$

If $\kappa > 1/2$, the bias term in (5.9) is dominated by the factor $\sqrt{n}$ and thus can be ignored; in this case, standard, conventional methods based on the Lindeberg-Feller CLT can be used to estimate confidence intervals for μ_λ. Otherwise, the bias is constant if $\kappa = 1/2$, or explodes if $\kappa < 1/2$. Note that $\kappa > 1/2$ if and only if $p+q \leq 2$ in the VRS case, or if and only if $p+q \leq 3$ in the CRS case. In the FDH case, this occurs only in the univariate case with $p = 1$, $q = 0$ or $p = 0$, $q = 1$. Replacing the scale factor $\sqrt{n}$ in (5.9) with n^γ, with $\gamma < \kappa \leq 1/2$, is not a viable option. Although doing so would make the bias disappear as $n \to \infty$, it would also cause the variance to converge to zero whenever $\kappa \leq 1/2$, making inference using the result in (5.9) impossible.

It is well-known that the nonparametric DEA and FDH estimators suffer from the curse of dimensionality, meaning that convergence rates become slower as $(p + q)$ increases. For purposes of estimating mean efficiency, the results of Kneip et al. (2015b) indicate the curse is even worse than before, with the "explosion" of bias coming at much smaller numbers of dimensions than found in many applied studies.

In general, whenever $\kappa \leq 1/2$, the results of Kneip et al. (2015b) make clear that conventional CLTs cannot be used to make inference about the mean μ_λ. The problem of controlling both bias and variance, for general number of dimensions $(p + q)$, can be addressed by using a different estimator of the population mean μ_λ and in addition rescaling the *estimator* of μ_λ by an appropriate factor different from $\sqrt{n}$ when $\kappa \leq 1/2$. Consider the factor $n_\kappa = \lfloor n^{2\kappa} \rfloor \leq n$, where $\lfloor a \rfloor$ denotes the integer part of a (note that this covers the limiting case of $\kappa = 1/2$). Then assume the observations in the sample $\mathcal{S}_n$ are randomly

ordered, and consider the latent estimator

$$\overline{\lambda}_{n_\kappa} = n_\kappa^{-1} \sum_{i=1}^{n_\kappa} \lambda(X_i, Y_i). \tag{5.10}$$

Of course, $\overline{\lambda}_{n_\kappa}$ is unobserved, but it can be estimated by

$$\widehat{\mu}_{n_\kappa} = n_\kappa^{-1} \sum_{i=1}^{n_\kappa} \widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n), \tag{5.11}$$

where the notation $\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n)$ serves to remind the reader that the individual efficiency estimates are computed from the full sample of n observations, while the sample mean is over $n_\kappa \leq n$ such estimates. Here again, one can use either the VRS, CRS or FDH version of the estimator.

Theorem 4.2 of Kneip et al. (2015b) establishes that when $\kappa \leq 1/2$,

$$n^\kappa \left(\widehat{\mu}_{n_\kappa} - \mu_\lambda - Cn^{-\kappa} - R_{n,\kappa} \right) \xrightarrow{\mathcal{L}} N(0, \sigma_\lambda^2) \tag{5.12}$$

as $n \to \infty$. Since $\sqrt{n_\kappa}\,(\widehat{\mu}_{n_\kappa} - \mu_\lambda)$ has a limiting distribution with unknown mean due to the bias term $Cn^{-\kappa}$, bootstrap approaches could be used to estimate the bias and hence to estimate confidence intervals for μ_λ. The variance could also be estimated by the same bootstrap, or by the consistent estimator

$$\widehat{\sigma}^2_{\lambda,n} = n^{-1} \sum_{i=1}^{n} \left(\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n) - \widehat{\mu}_n \right)^2. \tag{5.13}$$

Subsampling along the lines of Simar and Wilson (2011a) could also be used to make consistent inference about μ_λ. However, the estimator in (5.11) uses only a subset of the original n observations; unless n is extraordinarily large, taking subsamples among a subset of n_κ observations will likely leave too little information to provide useful inference.

Alternatively, Kneip et al. (2015b) demonstrate that the bias term $Cn^{-\kappa}$ can be estimated using a generalized jackknife estimator (e.g., see Gray and Schucany, 1972, Definition 2.1). Assume again that the observations (X_i, Y_i) are randomly ordered. Let $\mathcal{S}^{(1)}_{n/2}$ denote the set of the first $n/2$ observations in $\mathcal{S}_n$, and let $\mathcal{S}^{(2)}_{n/2}$ denote the set of remaining observations from $\mathcal{S}_n$. Let

$$\widehat{\mu}^*_{n/2} = \left(\widehat{\mu}^{(1)}_{n/2} + \widehat{\mu}^{(2)}_{n/2} \right) / 2. \tag{5.14}$$

Kneip et al. (2015b) show that

$$\begin{aligned}\widetilde{B}_{\kappa,n} &= (2^{\kappa}-1)^{-1}\left(\widehat{\mu}^{*}_{n/2}-\widehat{\mu}_{n}\right)\\ &= Cn^{-\kappa}+R_{n,\kappa}+o_{p}\left(n^{-1/2}\right)\end{aligned} \tag{5.15}$$

provides an estimator of the bias term $Cn^{-\kappa}$.

Of course, for n even there are $\binom{n}{n/2}$ possible splits of the sample $\mathcal{S}_n$. As noted by Kneip et al. (2016), the variation in $\widetilde{B}_{\kappa,n}$ can be reduced by repeating the above steps $K \ll \binom{n}{n/2}$ times, shuffling the observations before each split of $\mathcal{S}_n$, and then averaging the bias estimates. This yields a generalized jackknife estimate

$$\widehat{B}_{\kappa,n} = K^{-1}\sum_{k=1}^{K}\widetilde{B}_{\kappa,n,k}, \tag{5.16}$$

where $\widetilde{B}_{\kappa,n,k}$ represents the value computed from (5.15) using the kth sample split.

Theorem 4.3 of Kneip et al. (2015b) establishes that under appropriate assumptions, for $\kappa \geq 2/5$ for the VRS and CRS cases or $\kappa \geq 1/3$ for the FDH case,

$$\sqrt{n}\left(\widehat{\mu}_{n}-\widehat{B}_{\kappa,n}-\mu_{\lambda}+R_{n,\kappa}\right) \xrightarrow{\mathcal{L}} N(0,\sigma^{2}_{\lambda}). \tag{5.17}$$

as $n \to \infty$.

It is important to note that (5.17) is not valid for κ smaller that the bounds given in the theorem. This is due to the fact that for a particular definition of $R_{n,\kappa}$ (i.e., in either the VRS/CRS or FDH cases), values of κ smaller than the boundary value cause the remainder term, multiplied by $\sqrt{n}$, to diverge toward infinity. Interestingly, the normal approximation in (5.17) can be used with either the VRS-DEA or CRS-DEA estimators under the assumption of CRS if and only if $p+q \leq 5$; with the DEA-VRS estimator under convexity (but not CRS) if and only if $p+q \leq 4$; and with the FDH estimator assuming only free disposability (but not necessarily convexity nor CRS) if and only if $p+q \leq 3$. For these cases, an asymptotically correct $(1-\alpha)$ confidence interval for μ_λ is given by

$$\left[\widehat{\mu}_{n}-\widehat{B}_{\kappa,n} \pm z_{1-\alpha/2}\widehat{\sigma}_{\lambda,n}/\sqrt{n}\right], \tag{5.18}$$

where $z_{1-\alpha/2}$ is the corresponding quantile of the standard normal distribution and $\widehat{\sigma}_{\lambda,n}$ is given by (5.13).

In cases where κ is smaller than the bounds required by (5.17), the idea of estimating μ_λ by a sample mean of n_κ efficiency estimates as discussed above can be used with the bias estimate in 5.15, leading to Theorem 4.4 of Kneip et al. (2015b); i.e., under appropriate assumptions,

$$n^\kappa \left(\widehat{\mu}_{n_\kappa} - \widehat{B}_{\kappa,n} - \mu_\lambda + R_{n,\kappa} \right) \xrightarrow{\mathcal{L}} N(0, \sigma_\lambda^2). \tag{5.19}$$

as $n \to \infty$.

Equation 5.19 permits construction of consistent confidence intervals for μ_λ by replacing the unknown σ_λ^2 by its consistent estimator $\widehat{\sigma}_{\lambda,n}^2$. An asymptotically correct $1-\alpha$ confidence interval for μ_λ is given by

$$\left[\widehat{\mu}_{n_\kappa} - \widehat{B}_{\kappa,n} \pm z_{1-\alpha/2} \widehat{\sigma}_{\lambda,n} / n^\kappa \right], \tag{5.20}$$

where $z_{1-\alpha/2}$ is the corresponding quantile of the standard normal distribution. Here, the normal approximation can be used directly; bootstrap methods are not necessary.

Note that when $\kappa < 1/2$, the center of the confidence interval in (5.20) is determined by a random choice of $n_\kappa = n^{2\kappa} < n$ elements $\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n)$. This may be seen as arbitrary, but any confidence interval for μ_λ may be seen arbitrary in practice since asymmetric confidence intervals can be constructed by using different quantiles to establish the endpoints. The main point, however, is always to achieve a high level of coverage without making the confidence interval too wide to be informative.

Again for $\kappa < 1/2$, the arbitrariness of choosing a particular subsample of size n_κ in (5.20) can be eliminated by averaging the center of the interval in (5.20) over all possible draws (without replacement) of subsamples of size n_κ. Of course, this yields an interval centered on $\widehat{\mu}_n$, i.e.,

$$\left[\widehat{\mu}_n - \widehat{B}_{\kappa,n} \pm z_{1-\alpha/2} \widehat{\sigma}_{\lambda,n} / n^\kappa \right]. \tag{5.21}$$

The only difference between the intervals (5.20) and (5.21) is the centering value. Both intervals are equally informative, because they possess exactly the same length, $(2 z_{1-\alpha/2} \widehat{\sigma}_{\lambda,n} / n^\kappa)$. The interval (5.21) should be more accurate (i.e., should have higher coverage) because $\widehat{\mu}_n$ is a better estimator of μ_λ (i.e., has less mean-square error) than $\widehat{\mu}_{n_\kappa}$. If $\kappa < 1/2$, then $n_\kappa < n$, and hence the interval in (5.21) contains the true value μ_λ with probability greater than $1-\alpha$, since by the results above, it is clear that the coverage of the interval in (5.21) converges to 1 as $n \to \infty$. This is confirmed by the Monte Carlo evidence presented by Kneip et al. (2015b).

In cases with sufficiently small dimensions, either (5.17) or (5.19) can be used to provide different asymptotically valid confidence intervals for μ_λ. For the VRS-DEA and CRS-DEA estimators, this is possible whenever $\kappa = 2/5$ and so $n_\kappa < n$. The interval (5.18) uses the scaling $\sqrt{n}$ and neglects, in (5.17), a term $\sqrt{n}R_{n,\kappa} = O(n^{-1/10})$, whereas the interval (5.20) uses the scaling n^κ, neglecting in (5.19) a term $n^\kappa R_{n,\kappa} = O(n^{-1/5})$. We thus may expect a better approximation by using the interval (5.20). The same is true for the FDH case when $\kappa = 1/3$, where the interval (5.18) neglects terms of order $O(n^{-1/6})$ whereas the error when using (5.20) is only of order $O(n^{-1/3})$. These remarks are also confirmed by the Monte Carlo evidence reported by Kneip et al. (2015b).

5.2 Mean Conditional Efficiency

Daraio et al. (2018) extend the results of Kneip et al. (2015b) presented above in Section 5 to means of conditional efficiencies. Extension to the conditional case is complicated by the presence of the bandwidth h, which impacts convergence rates. Comparing (3.6) and (3.11) with the result in (4.17), it is apparent that the bandwidths in the conditional estimators reduce convergence rates from n^κ for the unconditional estimators to $n_h^\kappa = n^{\kappa/(\kappa r+1)}$ for the conditional estimators. Moreover, Theorem 4.1 of Daraio et al. (2018) establishes that under Assumptions 4.3–4.6, and in addition Assumption 4.12 for the FDH case or Assumptions 4.11 and 4.13 (referred to as "appropriate conditions" throughout the remainder of this subsection), as $n \to \infty$,

$$E\left(\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{X}_n) - \lambda^h(X_i, Y_i \mid Z_i)\right) = C_c n_h^{-\kappa} + R_{c,n_h,\kappa} \tag{5.22}$$

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

$$E\left(\left(\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{X}_n) - \lambda^h(X_i, Y_i \mid Z_i)\right)^2\right) = o\big(n_h^{-\kappa}\big), \tag{5.23}$$

and

$$\left|\mathrm{COV}\left(\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{X}_n) - \lambda^h(X_i, Y_i \mid Z_i),\ \widehat{\lambda}(X_j, Y_j \mid Z_j, \mathcal{X}_n) - \lambda^h(X_j, Y_j \mid Z_j)\right)\right| = o\big(n_h^{-1}\big) \tag{5.24}$$

for all $i, j \in \{1, \ldots, n\}$, $i \neq j$. In addition, for the conditional VRS-DEA estimator, $R_{c,n_h,\kappa} = O(n_h^{-3\kappa/2}(\log n_h)^{\alpha_1})$ while for the conditional FDH estimator $R_{c,n_h,\kappa} =$

$O(n_h^{-2\kappa}(\log n_h)^{\alpha_2})$. Note that incorporation of bandwidths in the conditional estimators reduces the order of the bias from Cn^{κ} to $C_c n_h^{\kappa}$.

Let

$$\widehat{\mu}_n = n^{-1} \sum_{i=1}^{n} \widehat{\lambda}(X_i, Y_i \mid \mathcal{X}_n) \tag{5.25}$$

and

$$\widehat{\mu}_{c,n} = n^{-1} \sum_{i=1}^{n} \widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{X}_n) \tag{5.26}$$

denote sample means of unconditional and conditional efficiency estimators. The efficiency estimators in (5.25) and (5.26) could be either FDH or VRS-DEA estimators; differences between the two are noted below when relevant. Next, define

$$\mu_c^h = E\left(\lambda^h(X, Y \mid Z)\right) = \int_{\mathcal{P}} \lambda^h(x, y \mid z) f_{XYZ}(x, y, z)\, dx\, dy\, dz \tag{5.27}$$

and

$$\sigma_c^{2,h} = \mathrm{VAR}\left(\lambda^h(X, Y \mid Z)\right) = \int_{\mathcal{P}} \left(\lambda^h(x, y \mid z) - \mu_c^h\right)^2 f_{XYZ}(x, y, z)\, dx\, dy\, dz, \tag{5.28}$$

where $\mathcal{P}$ is defined just before (4.1). These are the localized analogs of μ and σ^2. Next, let $\overline{\mu}_{c,n} = n^{-1} \sum_{i=1}^{n} \lambda^h(X_i, Y_i \mid Z_i)$. Although $\overline{\mu}_{c,n}$ is not observed, by the Lindeberg-Feller CLT

$$\frac{\sqrt{n}\left(\overline{\mu}_{c,n} - \mu_c^h\right)}{\sqrt{\sigma_c^{2,h}}} \xrightarrow{\mathcal{L}} N(0, 1) \tag{5.29}$$

under mild conditions.

Daraio et al. (2018) show that there can be no CLT for means of conditional efficiency estimators analogous to the result in (5.9) or Theorem 4.1 of Kneip et al. (2015b). There are no cases where standard CLTs with rate $n^{1/2}$ can be used with means of conditional efficiency estimators, unless Z is irrelevant with respect to the support of (X, Y), i.e., unless Assumption 4.1 holds. Given a sample of size n, there can be no CLT for means of conditional efficiency estimators based on a sample mean of all of the n conditional efficiency estimates.

Daraio et al. (2018) consider a random subsample $\mathcal{X}_{n_h}^*$ from $\mathcal{X}_n$ of size n_h where for simplicity we use the optimal rates for the bandwidths so that $n_h = O(n^{1/(\kappa r+1)})$. Define

$$\overline{\mu}_{c,n_h} = \frac{1}{n_h} \sum_{(X_i, Y_i, Z_i) \in \mathcal{X}_{n_h}^*} \lambda^h(X_i, Y_i \mid Z_i), \tag{5.30}$$

$$\widehat{\mu}_{c,n_h} = \frac{1}{n_h} \sum_{(X_i,Y_i,Z_i)\in\mathcal{X}^*_{n_h}} \widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{X}_n) \tag{5.31}$$

and let $\widetilde{\mu}_{c,n_h} = E(\widehat{\mu}_{c,n_h})$. Note that the estimators on the right-hand side (RHS) of (5.31) are computed relative to the full sample $\mathcal{X}_n$, but the summation is over elements of the sub-sample $\mathcal{X}^*_{n_h}$.

Theorem 4.2 of Daraio et al. (2018) establishes that

$$\sqrt{n_h}\left(\widehat{\mu}_{c,n_h} - \mu_c^h - C_c n_h^{-\kappa} - R_{c,n_h,\kappa}\right) / \sqrt{\sigma_c^{2,h}} \xrightarrow{\mathcal{L}} N(0,1), \tag{5.32}$$

and in addition establishes that $\sqrt{\sigma_c^{2,h}}$ is consistently estimated by

$$\widehat{\sigma}_{c,n}^{2,h} = n^{-1} \sum_{i=1}^{n} \left[\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{X}_n) - \widehat{\mu}_{c,n}\right]^2. \tag{5.33}$$

The result in (5.32) provides a CLT for means of conditional efficiency estimators, but the convergence rate is $\sqrt{n_h}$ as opposed to $\sqrt{n}$, and the result is of practical use only if $\kappa > 1/2$. If $\kappa = 1/2$, the bias term $C_c n_h^{-\kappa}$ does not vanish. If $\kappa < 1/2$, the bias term explodes as $n \to \infty$.

Similar to Kneip et al. (2015b), Daraio et al. (2018) propose a generalized jackknife estimator of the bias term $C_c n_n^{-\kappa}$. Assume the observations in $\mathcal{X}_n$ are randomly ordered. Let $\mathcal{X}^{(1)}_{n/2}$ denote the set of the first $n/2$ observations from $\mathcal{X}_n$, and let $\mathcal{X}^{(2)}_{n/2}$ denote the set of remaining $n/2$ observations from $\mathcal{X}_n$. Note that if n is odd, $\mathcal{X}^{(1)}_{n/2}$ can contain the first $\lfloor n/2 \rfloor$ observations and $\mathcal{X}^{(2)}_{n/2}$ can contain remaining $n - \lfloor n/2 \rfloor$ observations from $\mathcal{X}_n$. The fact that $\mathcal{X}^{(2)}_{n/2}$ contains one more observation than $\mathcal{X}^{(1)}_{n/2}$ makes no difference asymptotically. Next, for $j \in \{1,\ 2\}$ define

$$\widehat{\mu}^j_{c,n/2} = (n/2)^{-1} \sum_{(X_i,Y_i,Z_i)\in\mathcal{X}^{(j)}_{n/2}} \widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{X}^{(j)}_{n/2}). \tag{5.34}$$

Now define

$$\widehat{\mu}^*_{c,n/2} = \left(\widehat{\mu}^1_{c,n/2} + \widehat{\mu}^2_{c,n/2}\right)/2. \tag{5.35}$$

Then

$$\widetilde{B}^c_{\kappa,n_h} = (2^\kappa - 1)^{-1}\left(\widehat{\mu}^*_{c,n/2} - \widehat{\mu}_{c,n}\right) \tag{5.36}$$

is an estimator of the leading bias term $C_c n_h^{-\kappa}$ in (5.32). Averaging over $K \ll \binom{n}{n/2}$ splits of the sample $\mathcal{X}_n$ as before yields the generalized jackknife estimator

$$\widehat{B}^c_{\kappa,n_h} = K^{-1} \sum_{k=1}^{K} \widetilde{B}^c_{\kappa,n_h,k}, \tag{5.37}$$

where $\widetilde{B}^c_{\kappa,n_h,k}$ represents the value computed from (5.36) using the kth sample split.

Theorem 4.3 of Daraio et al. (2018) establishes that under appropriate conditions, with $\kappa = 1/(p+q) \geq 1/3$ in the FDH case or $\kappa = 2/(p+q+1) \geq 2/5$ in the VRS-DEA case,

$$\frac{\sqrt{n_h}\left(\widehat{\mu}_{c,n_h} - \mu^h_c - \widehat{B}^c_{\kappa,n_h} - R^*_{c,n_h,\kappa}\right)}{\sqrt{\sigma^{2,h}_c}} \xrightarrow{\mathcal{L}} N(0,1) \tag{5.38}$$

as $n \to \infty$. Alternatively, Theorem 4.4 of Daraio et al. (2018) establishes that under appropriate conditions, whenever $\kappa < 1/2$,

$$\frac{\sqrt{n_{h,\kappa}}\left(\widehat{\mu}_{c,n_{h,\kappa}} - \mu^h_c - \widehat{B}^c_{\kappa,n_h} - R^*_{c,n_h,\kappa}\right)}{\sqrt{\sigma^{2,h}_c}} \xrightarrow{\mathcal{L}} N(0,1), \tag{5.39}$$

as $n \to \infty$. The results in (5.38) and (5.39) provide CLTs for means of conditional efficiencies covering all values of κ, and can be used to estimate confidence intervals or for testing Assumption 4.1 versus Assumption 4.2 as described below in Section 6.4.

6 Hypothesis Testing

6.1 Testing Convexity versus Non-Convexity

In situations where one might want to test convexity of Ψ versus non-convexity, typically a single iid sample $\mathcal{S}_n = \{(X_i, Y_i\}_{i=1}^n$ of n input-output pairs is available. Under the null hypothesis of convexity, both the FDH and VRS-DEA estimators are consistent, but under the alternative, only the FDH estimator is consistent. It might be tempting to compute the sample means

$$\widehat{\mu}^{\text{full}}_{\text{VRS},n} = n^{-1} \sum_{i=1}^{n} \widehat{\lambda}_{\text{VRS}}\left(X_i, Y_i \mid \mathcal{S}_n\right) \tag{6.1}$$

and

$$\widehat{\mu}^{\text{full}}_{\text{FDH},n} = n^{-1} \sum_{(X_i,Y_i)\in\mathcal{S}_n} \widehat{\lambda}_{\text{FDH}}\left(X_i, Y_i \mid \mathcal{S}_n\right) \tag{6.2}$$

using the full set of observations in $\mathcal{S}_n$ and use this with (6.1) to construct a test statistic based on the difference $\widehat{\mu}^{\text{full}}_{\text{FDH},n} - \widehat{\mu}^{\text{full}}_{\text{VRS},n}$. By construction, $\widehat{\lambda}_{\text{VRS}}(X_i, Y_i \mid \mathcal{S}_n) \geq \widehat{\lambda}_{\text{FDH}}(X_i, Y_i \mid \mathcal{S}_n) \geq 1$ and therefore $\widehat{\mu}^{\text{full}}_{\text{VRS},n} - \widehat{\mu}^{\text{full}}_{\text{FDH},n} \geq 0$. Under the null, $\widehat{\mu}^{\text{full}}_{\text{FDH},n} - \widehat{\mu}^{\text{full}}_{\text{VRS},n}$ is expected to be "small," while under the alternative the difference is expected to be "large."

Unfortunately, such an approach is doomed to failure. Using the output-oriented analog of Theorem 4.1 of Kneip et al. (2015b) and reasoning similar to that of Kneip et al. (2016, Section 3.2), it can be shown that $n^a \left(\widehat{\lambda}_{\text{VRS}}(X_i, Y_i \mid \mathcal{S}_n) - \widehat{\lambda}_{\text{FDH}}(X_i, Y_i \mid \mathcal{S}_n) \right)$ converges under the null to a degenerate distribution for any power $a \leq 1/2$ of n; i.e., the asymptotic variance of the statistic is zero, and the density of the statistic converges to a Dirac delta function at zero under the null, rendering inference impossible.

Kneip et al. (2016) solve this problem by randomly splitting the sample $\mathcal{S}_n$ into two mutually exclusive, collectively exhaustive parts $\mathcal{S}_{1,n_1}$ and $\mathcal{S}_{2,n_2}$ such that $\mathcal{S}_{1,n_1} \cap \mathcal{S}_{2,n_2} = \emptyset$ and $\mathcal{S}_{1,n_1} \cup \mathcal{S}_{2,n_2} = \mathcal{S}_n$. Recall that the FDH estimator converges at rate $n^{1/(p+q)}$, while the VRS-DEA estimator converges at rate $n^{2/(p+q+1)}$ under VRS or at rate $n^{2/(p+q)}$ under CRS. Kneip et al. (2016) suggest exploiting this difference by setting $n_1^{2/(p+q+1)} = \beta n_2^{1/(p+q)}$ and $n_1 + n_2 = n$ for a given sample size n, where β is a constant, and then solving for n_1 and n_2. There is no closed-form solution, but it is easy to find a numerical solution by writing $n - n_1 - \beta^{-1} n_1^{2(p+q)/(p+q+1)} = 0$; the root of this equation is bounded between 0 and $n/2$, and can be found by simple bisection. Letting n_1 equal the integer part of the solution and setting $n_2 = n - n_1$ gives the desired subsample sizes with $n_2 > n_1$. Using the larger subsample $\mathcal{S}_{2,n_2}$ to compute the FDH estimates and the smaller subsample $\mathcal{S}_{1,n_1}$ to compute the VRS-DEA estimates allocates observations from the original sample $\mathcal{S}_n$ efficiently in the sense that more observations are used to mitigate the slower convergence rate of the FDH estimator. Simulation results provided by Kneip et al. (2016) suggest that the choice of value for β matters less as sample size n increases, and that setting $\beta = 1$ gives reasonable results across various values of n and $(p + q)$.

Once the original sample has been split, the sample means

$$\widehat{\mu}_{\text{VRS},n_1} = n_1^{-1} \sum_{(X_i,Y_i) \in \mathcal{S}_{1,n_1}} \widehat{\lambda}_{\text{VRS}}(X_i, Y_i \mid \mathcal{S}_{1,n_1}) \tag{6.3}$$

and

$$\widehat{\mu}_{\text{FDH},n_2} = n_2^{-1} \sum_{(X_i,Y_i) \in \mathcal{S}_{2,n_2}} \widehat{\lambda}_{\text{FDH}}\left(X_i, Y_i \mid \mathcal{S}_{2,n_2}\right) \tag{6.4}$$

can be computed. In addition, the sample variances

$$\widehat{\sigma}^2_{\text{VRS},n_2} = n_2^{-1} \sum_{(X_i,Y_i) \in \mathcal{S}_{2,n_2}} \left[\widehat{\lambda}_{\text{VRS}}(X_i, Y_i \mid \mathcal{S}_{2,n_2}) - \widehat{\mu}_{\text{VRS},n_2} \right]^2. \tag{6.5}$$

and

$$\widehat{\sigma}^2_{\mathrm{FDH},n_2} = n_2^{-1} \sum_{(X_i,Y_i)\in\mathcal{S}_{2,n_2}} \Big[\widehat{\lambda}_{\mathrm{FDH}}(X_i,Y_i \mid \mathcal{S}_{2,n_2}) - \widehat{\mu}_{\mathrm{FDH},n_2}\Big]^2. \tag{6.6}$$

can also be computed. Now let $\kappa_1 = 2/(p+q+1)$ and $\kappa_2 = 1/(p+q)$. Then the corresponding bias estimates $\widehat{B}_{\mathrm{VRS},\kappa_1,n_1}$ and $\widehat{B}_{\mathrm{FDH},\kappa_2,n_2}$ can also be computed from the two parts of the full sample, which requires further splitting both of the two parts $\mathcal{S}_{1,n_1}$ and $\mathcal{S}_{2,n_2}$ along the lines discussed above in Section 5.1.

The rate of the FDH estimator is slower than the rate of the VRS-DEA estimator, and hence the rate of the FDH estimator dominates. Kneip et al. (2016) show that for $(p+q) \leq 3$, the test statistic

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

can be used to test the null hypothesis of convexity for Ψ versus the alternative hypothesis that Ψ is not convex.

Alternatively, if $(p+q) > 3$, the sample means must be computed using subsets of $\mathcal{S}_{1,n_1}$ and $\mathcal{S}_{2,n_2}$. For $\ell \in \{1,2\}$, let $\kappa = \kappa_2 = 1/(p+q)$ and for $\ell \in \{1,\ 2\}$ let $n_{\ell,\kappa} = \lfloor n_\ell^{2\kappa} \rfloor$ so that $n_{\ell,\kappa} < n_\ell$ for $\kappa < 1/2$. Let $\mathcal{S}_{\ell,n_{\ell,\kappa}}$ be a random subset of $n_{\ell,\kappa}$ input-output pairs from $\mathcal{S}_{\ell,n_\ell}$. Then let

$$\widehat{\mu}_{\mathrm{VRS},n_{1,\kappa}} = n_{1,\kappa}^{-1} \sum_{(X_i,Y_i)\in\mathcal{S}^*_{1,n_{1,\kappa}}} \widehat{\lambda}_{\mathrm{VRS}}(X_i,Y_i \mid \mathcal{S}_{1,n_1}) \tag{6.8}$$

and

$$\widehat{\mu}_{\mathrm{FDH},n_{2,\kappa}} = n_{2,\kappa}^{-1} \sum_{(X_i,Y_i)\in\mathcal{S}^*_{2,n_{2,\kappa}}} \widehat{\lambda}_{\mathrm{FDH}}(X_i,Y_i \mid \mathcal{S}_{2,n_2}), \tag{6.9}$$

noting that the summations in (6.8) and (6.9) are over subsets of the observations used to compute the efficiency estimates under the summation signs.

Then by Theorem 4.4 of Kneip et al. (2015b), for $(p+q) > 3$,

$$\widehat{\tau}_{2,n} = \frac{\big(\widehat{\mu}_{\mathrm{VRS},n_{1,\kappa}} - \widehat{\mu}_{\mathrm{FDH},n_{2,\kappa}}\big) - \Big(\widehat{B}_{\mathrm{VRS},\kappa_1,n_1} - \widehat{B}_{\mathrm{FDH},\kappa_2,n_2}\Big)}{\sqrt{\frac{\widehat{\sigma}^2_{\mathrm{VRS},n_1}}{n_{1,\kappa}} + \frac{\widehat{\sigma}^2_{\mathrm{FDH},n_2}}{n_{2,\kappa}}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{6.10}$$

under the null hypothesis of convexity for Ψ. Note that $\widehat{\tau}_{2,n}$ differs from $\widehat{\tau}_{1,n}$ both in terms of the number of efficiency estimates used in the sample means as well as the divisors of the variance terms under the square-root sign.

Depending on whether $(p+q) \leq 3$ or $(p+q) > 3$, either $\widehat{\tau}_{1,n}$ or $\widehat{\tau}_{2,n}$ can be used to test the null hypothesis of constant returns to scale, with critical values obtained from the standard normal distribution. In particular, for $j \in \{1, 2\}$, the null hypothesis of convexity of Ψ is rejected if $\widehat{p} = 1 - \Phi(\widehat{\tau}_{j,n})$ is less than a suitably small value, e.g., .1, .05, or .01.

6.2 Testing Constant versus Variable Returns to Scale

Kneip et al. (2016) use the CLT results of Kneip et al. (2015b) discussed above in Section 5.1 to develop a test of CRS versus VRS for the technology Ψ^{∂}. A key result is that under CRS, the VRS-DEA estimator attains same convergence rate $n^{2/)(p+q)}$ as the CRS estimator as established by Theorem 3.1 of Kneip et al. (2016).

Under the null hypothesis of CRS, both the VRS-DEA and CRS-DEA estimators of $\lambda(X, Y)$ are consistent, but under the alternative, only the VRS-DEA estimator is consistent. Consider the sample means given by (6.1) and

$$\widehat{\mu}^{\text{full}}_{\text{CRS},n} = n^{-1} \sum_{i=1}^{n} \widehat{\lambda}_{\text{CRS}} \left(X_i, Y_i \mid \mathcal{S}_n \right) \tag{6.11}$$

computed using all of the n observations in $\mathcal{S}_n$. Under the null, one would expect $\widehat{\mu}^{\text{full}}_{\text{CRS},n} - \widehat{\mu}^{\text{full}}_{\text{VRS},n}$ to be "small," while under the alternative $\widehat{\mu}^{\text{full}}_{\text{CRS},n} - \widehat{\mu}^{\text{full}}_{\text{VRS},n}$ is expected to be "large." However, as shown by Kneip et al. (2016), a test statistic using the difference in the sample means given by (6.1)–(6.11) will have a degenerate distribution under the null since the asymptotic variance of $\left(\widehat{\mu}^{\text{full}}_{\text{CRS},n} - \widehat{\mu}^{\text{full}}_{\text{VRS},n}\right)$ is zero, similar to the case of testing convexity versus non-convexity discussed above in Section 6.1. Consequently, sub-sampling methods are needed here just as they were in Section 6.1.

In order to obtain non-degenerate test statistics, randomly split the sample into two samples $\mathcal{S}_{1,n_1}$, $\mathcal{S}_{2,n_2}$ such that $\mathcal{S}_{1,n_1} \cup \mathcal{S}_{2,n_2} = \mathcal{S}_n$ and $\mathcal{S}_{1,n_1} \cap \mathcal{S}_{2,n_2} = \emptyset$, where $n_1 = \lfloor n/2 \rfloor$ and $n_2 = n - n_1$. Next, let

$$\widehat{\mu}_{\text{CRS},n_2} = n_2^{-1} \sum_{(X_i,Y_i) \in \mathcal{S}_{2,n_2}} \widehat{\lambda}_{\text{CRS}}(X_i, Y_i \mid \mathcal{S}_{2,n_2}). \tag{6.12}$$

and

$$\widehat{\sigma}^2_{\text{CRS},n_2} = n_2^{-1} \sum_{(X_i,Y_i) \in \mathcal{S}_{2,n_2}} \left[\widehat{\lambda}_{\text{CRS}}(X_i, Y_i \mid \mathcal{S}_{2,n_2}) - \widehat{\mu}_{\text{CRS},n_2} \right]^2, \tag{6.13}$$

and recall (6.3) and (6.5) from the discussion in Section 6.1.

The (partial) sample $\mathcal{S}_{2,n_2}$ can be used to compute an estimate $\widehat{B}_{\mathrm{CRS},\kappa,n_2}$ of bias for the CRS estimator by splitting $\mathcal{S}_{2,n_2}$ into two parts along the lines discussed above. Then under the null hypothesis of constant returns to scale, Kneip et al. (2016) demonstrate that

$$\widehat{\tau}_{3,n} = \frac{(\widehat{\mu}_{\mathrm{CRS},n_2} - \widehat{\mu}_{\mathrm{VRS},n_1}) - \left(\widehat{B}_{\mathrm{CRS},\kappa,n_2} - \widehat{B}_{\mathrm{VRS},\kappa,n_1}\right)}{\sqrt{\frac{\widehat{\sigma}^2_{\mathrm{VRS},n_1}}{n_1} + \frac{\widehat{\sigma}^2_{\mathrm{CRS},n_2}}{n_2}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{6.14}$$

provided $(p+q) \leq 5$.

Alternatively, if $(p+q) > 5$, the sample means must be computed using subsets of the available observations. For $\ell \in \{1,2\}$ and $\mathcal{S}^*_{\ell,n_{\ell,\kappa}}$ defined as in Section 6.1, let

$$\widehat{\mu}_{\mathrm{CRS},n_{2,\kappa}} = n^{-1}_{2,\kappa} \sum_{(X_i,Y_i)\in\mathcal{S}^*_{2,n_{2,\kappa}}} \widehat{\lambda}_{\mathrm{CRS}}(X_i, Y_i \mid \mathcal{S}_{2,n_2}). \tag{6.15}$$

and recall the definition of $\widehat{\mu}_{\mathrm{VRS},n_{1,\kappa}}$ in (6.8). Here again, the summation in (6.15) is over a random subset of the observations used to compute the efficiency estimates under the summation sign. Again under the null hypothesis of constant returns to scale, by Theorem 4.4 of Kneip et al. (2015b), and Theorem 3.1 of Kneip et al. (2016),

$$\widehat{\tau}_{4,n} = \frac{\left(\widehat{\mu}_{\mathrm{CRS},n_{2,\kappa}} - \widehat{\mu}_{\mathrm{VRS},n_{1,\kappa}}\right) - \left(\widehat{B}_{\mathrm{CRS},\kappa,n_2} - \widehat{B}_{\mathrm{VRS},\kappa,n_1}\right)}{\sqrt{\frac{\widehat{\sigma}^2_{\mathrm{VRS},n_1}}{n_{1,\kappa}} + \frac{\widehat{\sigma}^2_{\mathrm{CRS},n_2}}{n_{2,\kappa}}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{6.16}$$

for $(p+q) > 5$. Similar to the comparison between $\widehat{\tau}_{2,n}$ and $\widehat{\tau}_{1,n}$ in Section 6.1, $\widehat{\tau}_{4,n}$ differs from $\widehat{\tau}_{3,n}$ both in terms of the number of efficiency estimates used to compute the sample means in the numerator as well as the divisors of the variance terms under the square-root sign in the denominator.

Depending on the value of $(p+q)$, either $\widehat{\tau}_{3,n}$ or $\widehat{\tau}_{4,n}$ can be used to test the null hypothesis of constant returns to scale, with critical values obtained from the standard normal distribution. In particular, for $j \in \{3,\ 4\}$, the null hypothesis of constant returns to scale is rejected if $\widehat{p} = 1 - \Phi(\widehat{\tau}_{j,n})$ is less than, say, .1, .05, or .01.

6.3 Testing for Differences in Mean Efficiency Across Groups of Producers

Testing for differences in mean efficiency across two groups was suggested—but not implemented—in the application of Charnes et al. (1981), who considered two groups of

schools, one receiving a treatment effect and the other not receiving the treatment. There are many situations where such a test might be useful. For example, one might test whether mean efficiency among for-profit producers is greater than mean efficiency of non-profit producers in studies of hospitals, banks and credit unions, or perhaps other industries. One might similarly be interested in comparing average performance of publicly-traded versus privately-held firms, or in regional differences that might reflect variation in state-level regulation or other industry features. However, the discussion above in Section 5 makes clear that standard CLT results are not useful in general when considering means of nonparametric efficiency estimates.

Consider two group of firms G_1 and G_2 of sizes n_1 and n_2. Suppose the researcher wishes to test whether $\mu_{1,\lambda} = E(\lambda(X,Y) \mid (X,Y) \in G_1)$ and $\mu_{2,\lambda} = E(\lambda(X,Y) \mid (X,Y) \in G_2)$ are equal against the alternative that Group 1 has greater mean efficiency. More formally, one might test the null hypothesis $H_0\colon \mu_{1,\lambda} = \mu_{2,\lambda}$ versus the alternative hypothesis $H_1\colon \mu_{1,\lambda} > \mu_{2,\lambda}$. One could also conduct a test with a two-sided alternative; such a test would of course follow a procedure similar to the one outlined here for a one-sided test.

The test discussed here makes no restriction on whether firms in the two groups face the same frontier, i.e., whether they operate in the same production set. Kneip et al. (2016, Section 3.1.2) discuss a version of the test where under the null firms in the two groups face the same frontier. Readers can consult Kneip et al. (2016) for details on the alternative form of the test.

Suppose iid samples $\mathcal{S}_{1,n_1} = \{(X_i,Y_i)\}_{i=1}^{n_1}$ and $\mathcal{S}_{2,n_2} = \{(X_i,Y_i)\}_{i=1}^{n_2}$ of input-output pairs from groups 1 and 2 (respectively) are available. In addition, assume these samples are independent of each other. The test here is simpler than the tests of convexity versus non-convexity and constant versus variable returns to scale since two independent samples are available initially. The two samples yield independent estimators

$$\widehat{\mu}_{1,n_1} = n_1^{-1} \sum_{(X_i,Y_i)\in\mathcal{S}_{1,n_1}} \widehat{\lambda}(X_i,Y_i \mid \mathcal{S}_{1,n_1}) \tag{6.17}$$

and

$$\widehat{\mu}_{2,n_2} = n_2^{-1} \sum_{(X_i,Y_i)\in\mathcal{S}_{2,n_2}} \widehat{\lambda}(X_i,Y_i \mid \mathcal{S}_{2,n_2}) \tag{6.18}$$

of $\mu_{1,\lambda}$ and $\mu_{2,\lambda}$, respectively; the conditioning indicates the sample used to compute the efficiency estimates under the summation signs. As in Sections 6.1 and 6.2, the subscripts

on $\widehat{\lambda}(\cdot)$ have been dropped; either the FDH, VRS-DEA, or CRS-DEA estimators with corresponding convergence rates n^{κ} could be used, although the same estimator would be used for both groups. Theorem 4.1 of Kneip et al. (2015b) establishes (under appropriate regularity conditions; see Section 5.1 for details) consistency and other properties of these estimators. The same theorem, however, makes clear that standard, conventional central limit theorems can be used to make inference about the population means $\mu_{1,\lambda}$ and $\mu_{2,\lambda}$ only when the dimensionality $(p+q)$ is small enough so that $\kappa > 1/2$ due to the bias of the estimators $\widehat{\mu}_{1,n_1}$ and $\widehat{\mu}_{2,n_2}$.

As discussed earlier in Section 5, the Lindeberg-Feller and other central limit theorems fail when FDH, VRS-DEA, or CRS-DEA estimators are averaged as in (6.17)–(6.18) due to the fact that while averaging drives the variance to zero, it does not diminish the bias. From Kneip et al. (2015b) and the discussion in Section 5.1, it can be seen that unless $(p+q)$ is very small, scaling sample means such as (6.17)–(6.18) by a power of the sample size to stabilize the bias results in a degenerate statistic with the variance converging to zero. On the other hand, scaling if the power of the sample size is chosen to stabilize the variance, the bias explodes. Hence the bias must be estimated and dealt with.

For each of the two groups $\ell \in \{1,\ 2\}$, a bias estimate $\widehat{B}_{\ell,\kappa,n_\ell}$ can be obtained as described earlier in Section 5.1. Then using using VRS-DEA estimators with $p+q \leq 4$ (or CRS-DEA estimators with $p+q \leq 5$, or FDH estimators with $p+q \leq 3$), the test statistic

$$\widehat{\tau}_{5,n_1,n_2} = \frac{(\widehat{\mu}_{1,n_1} - \widehat{\mu}_{2,n_2}) - \left(\widehat{B}_{1,\kappa,n_1} - \widehat{B}_{2,\kappa,n_2}\right) - (\mu_{1,\lambda} - \mu_{2,\lambda})}{\sqrt{\frac{\widehat{\sigma}^2_{1,n_1}}{n_1} + \frac{\widehat{\sigma}^2_{2,n_2}}{n_2}}} \xrightarrow{\mathcal{L}} N(0,1), \tag{6.19}$$

can be used to test the null hypothesis of equivalent mean efficiency for groups 1 and 2 provided $n_1/n_2 \to c > 0$ as $n_1, n_2 \to \infty$, where c is a constant. The variance estimates appearing in the denominator of $\widehat{\tau}_{5,n_1,n_2}$ are given by

$$\widehat{\sigma}^2_{\ell,n_\ell} = n_\ell^{-1} \sum_{i=1}^{n_\ell} \left[\widehat{\lambda}(X_{\ell,i}, Y_{\ell,i} \mid \mathcal{S}_{\ell,n_\ell}) - \widehat{\mu}_{\ell,n_\ell}\right]^2 \xrightarrow{p} \sigma^2_\ell \tag{6.20}$$

for $\ell \in \{1,2\}$; and all values of $(p+q)$.

In situations where $p+q > 4$ with VRS-DEA estimators (or $p+q > 5$ with CRS-DEA estimators, or $p+q > 3$ with FDH estimators), means based on sub-samples must be used. For $\ell \in \{1,2\}$, let $n_{\ell,\kappa} = \lfloor n_\ell^{2\kappa} \rfloor$; then $n_{\ell,\kappa} < n_\ell$ for $\kappa < 1/2$. Let $\mathcal{S}^*_{\ell,n_{\ell,\kappa}}$ be a random subset of

$n_{\ell,\kappa}$ input-output pairs from $\mathcal{S}_{\ell,n_\ell}$. Then let

$$\widehat{\mu}_{\ell,n_{\ell,\kappa}} = n_{\ell,\kappa}^{-1} \sum_{(X_{\ell,i},Y_{\ell,i})\in\mathcal{S}^*_{\ell,n_{\ell,\kappa}}} \widehat{\lambda}\left(X_{\ell,i},Y_{\ell,i} \mid \mathcal{S}_{\ell,n_\ell}\right), \tag{6.21}$$

noting that while the summation is over only the input-output pairs in $\mathcal{S}^*_{\ell,n_{\ell,\kappa}}$, the efficiency estimates under the summation sign are computed using all of the input-output pairs in $\mathcal{S}_{\ell,n_\ell}$.

Kneip et al. (2016) obtain a similar test statistic for use in situations where $p+q>4$ with VRS-DEA estimators, $p+q>5$ with CRS-DEA estimators, or $p+q>3$ with FDH estimators. In particular,

$$\widehat{\tau}_{6,n_{1,\kappa},n_{2,\kappa}} = \frac{\left(\widehat{\mu}_{1,n_{1,\kappa}} - \widehat{\mu}_{2,n_{2,\kappa}}\right) - \left(\widehat{B}_{1,\kappa,n_1} - \widehat{B}_{2,\kappa,n_2}\right) - (\mu_{1,\lambda}-\mu_{2,\lambda})}{\sqrt{\frac{\widehat{\sigma}^2_{1,n_1}}{n_{1,\kappa}} + \frac{\widehat{\sigma}^2_{2,n_2}}{n_{2,\kappa}}}} \xrightarrow{\mathcal{L}} N(0,1), \tag{6.22}$$

again provided $n_1/n_2 \to c>0$ as $n_1, n_2 \to \infty$. Note that the same estimates for the variances and biases are used in (6.22) as in (6.19). The only difference between (6.19) and (6.22) is in the number of observations used to compute the sample means and in the quantities that divide the variance terms under the square-root sign in the denominator.

Kneip et al. (2016) note that the results in (6.19) and (6.22) hold for *any* values of $\mu_{1,\lambda}$ and $\mu_{2,\lambda}$. Hence, if one tests $H_0\colon \mu_{1,\lambda} = \mu_{2,\lambda}$ versus an alternative hypothesis such as $H_1\colon \mu_{1,\lambda} > \mu_{2,\lambda}$ or perhaps $H_1\colon \mu_{1,\lambda} \neq \mu_{2,\lambda}$, the (asymptotic) distribution of the test statistic will be known under the null and up to $\mu_{1,\lambda},\ \mu_{2,\lambda}$ under the alternative hypothesis. Consequently, given two independent samples, one can either (i) compute under the null (so that $(\mu_{1\lambda}-\mu_{2,\lambda})=0$) $\widehat{\tau}_{5,n_1,n_2}$ or $\widehat{\tau}_{6,n_{1,\kappa},n_{2,\kappa}}$ as appropriate, and compare the resulting value against a critical value from the $N(0,1)$ distribution, or (ii) use (6.19) and (6.22) to estimate a confidence interval for $(\mu_{1\lambda}-\mu_{2,\lambda})$. If the estimated interval does not include 0, one would reject the null; otherwise, one would fail to reject the null. Furthermore, the outcome will be the same under either approach; i.e., for a given test size, both approaches will either reject or fail to reject the null. It clearly follows that for a given departure from the null, the tests will reject the null with probability tending to one as $n_1, n_2 \to \infty$, and hence the tests are consistent.

6.4 Testing Separability

Daraio et al. (2018) develop a test of separability (Assumption 4.1) versus the alternative of non-separability (Assumption 4.2) using the CLT results for both unconditional and conditional efficiencies discussed above in Section 5. The idea for building a test statistic is to compare the conditional and unconditional efficiency scores using relevant statistics that are functions of $\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n)$ and $\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_n)$ for $i = 1, \ldots, n$. Note that under Assumption 4.1, $\lambda(X, Y) = \lambda(X, Y \mid Z)$ with probability one, even if Z may influence the distribution of the inefficiencies inside the attainable set, and the two estimators converge to the same object. But under Assumption 4.2, the conditional attainable sets Ψ^z are different and the two estimators converge to different objects. Moreover, under Assumption 4.2, $\lambda(X, Y) \geq \lambda(X, Y \mid Z)$ with strict inequality holding for some $(X, Y, Z) \in \mathcal{P}$.

In order to implement the test of separability, randomly split the sample $\mathcal{S}_n$ into two independent, parts $\mathcal{S}_{1,n_1}$, $\mathcal{S}_{2,n_2}$ such that $n_1 = \lfloor n/2 \rfloor$, $n_2 = n - n_1$, $\mathcal{S}_{1,n_1} \bigcup \mathcal{S}_{2,n_2} = \mathcal{S}_n$, and $\mathcal{S}_{1,n_1} \bigcap \mathcal{S}_{2,n_2} = \emptyset$. The n_1 observations in $\mathcal{S}_{1,n_1}$ are used for the unconditional estimates, while the n_2 observations in $\mathcal{S}_{2,n_2}$ are used for the conditional estimates.[13]

After splitting the sample, compute the estimators

$$\widehat{\mu}_{n_1} = n_1^{-1} \sum_{(X_i, Y_i) \in \mathcal{S}_{1,n_1}} \widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_{1,n_1}) \tag{6.23}$$

and

$$\widehat{\mu}_{c,n_{2,h}} = n_{2,h}^{-1} \sum_{(X_i, Y_i, Z_i) \in \mathcal{S}^*_{2,n_{2,h}}} \widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_{2,n_2}), \tag{6.24}$$

where as above in Section 5, $\mathcal{S}^*_{2,n_{2,h}}$ in (6.24), is a random subsample from $\mathcal{S}_{2,n_2}$ of size $n_{2,h} = \min(n_2, n_2 h^r)$. Consistent estimators of the variances are given in the two independent

[13] Kneip et al. (2016) proposed splitting the sample unevenly to account for the difference in the convergence rates between the (unconditional) DEA and FDH estimators used in their convexity test, giving more observations to the subsample used to compute FDH estimates than to the subsample used to compute DEA estimates. Recall that the unconditional efficiency estimators converge at rate n^{κ}, while the conditional efficiency estimators converge at rate n_h^{κ}. The optimal bandwidths are of order $n^{-\kappa/(r\kappa+1)}$, giving a rate of $n^{\kappa/(r\kappa+1)}$ for the conditional efficiency estimators. Using the logic of Kneip et al. (2016), the full sample $\mathcal{S}_n$ can be split so that the estimators in the two subsamples achieve the same rate of convergence by setting $n_1^{\kappa} = n_2^{\kappa/(r\kappa+1)}$. This gives $n_1 = n_2^{1/(r\kappa+1)}$. Values of n_1, n_2 are obtained by finding the root η_0 in $n - \eta - \eta^{1/(r\kappa+1)} = 0$ and setting $n_2 = \lfloor \eta_0 \rfloor$ and $n_1 = n - n_2$. However, this will often result in too few observations in the first subsample to obtain meaningful results. For example, if $p = q = r = 1$ and $n = 200$, following the reasoning above would lead to $n_1 = 22$ and $n_2 = 178$.

samples by

$$\widehat{\sigma}^2_{n_1} = n_1^{-1} \sum_{(X_i,Y_i)\in\mathcal{S}_{1,n_1}} \left(\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_{1,n_1}) - \widehat{\mu}_{n_1}\right)^2 \tag{6.25}$$

and

$$\widehat{\sigma}^{2,h}_{c,n_2} = n_2^{-1} \sum_{(X_i,Y_i,Z_i)\in\mathcal{S}_{2,n_2}} \left(\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_{2,n_2}) - \widehat{\mu}_{c,n_2}\right)^2 \tag{6.26}$$

(respectively), where the full (sub)samples are used to estimate the variances.

The estimators of bias for a single split of each subsample for the unconditional and conditional cases are given by $\widehat{B}_{\kappa,n_1}$ for the unconditional case and $\widehat{B}^c_{\kappa,n_{2,h}}$ for the conditional case as described in Sections 5.1 and 5.2, respectively.

For values of $(p+q)$ such that $\kappa \geq 1/3$ in the FDH case or $\kappa \geq 2/5$ when DEA estimators are used, the CLT results in (5.17) and (5.38) can be used to construct an asymptotically normal test statistic for testing the null hypothesis of separability. Since the bias-corrected sample means are independent due to splitting the original sample into independent parts, and since two sequences of independent variables each with normal limiting distributions have a joint bivariate normal limiting distribution with independent marginals, if follows that for the values of $(p+q)$ given above

$$\tau_{7,n} = \frac{\left(\widehat{\mu}_{n_1} - \widehat{\mu}_{c,n_{2,h}}\right) - \left(\widehat{B}_{\kappa,n_1} - \widehat{B}^c_{\kappa,n_{2,h}}\right)}{\sqrt{\frac{\widehat{\sigma}^2_{n_1}}{n_1} + \frac{\widehat{\sigma}^{2,h}_{c,n_2}}{n_{2,h}}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{6.27}$$

under the null. Alternatively, for $\kappa < 1/2$, similar reasoning and using the CLT results in (5.19) and (5.39) leads to

$$\tau_{8,n} = \frac{\left(\widehat{\mu}_{n_{1,\kappa}} - \widehat{\mu}_{c,n_{2,h,\kappa}}\right) - \left(\widehat{B}_{\kappa,n_1} - \widehat{B}^c_{\kappa,n_{2,h}}\right)}{\sqrt{\frac{\widehat{\sigma}^2_{n_1}}{n_{1,\kappa}} + \frac{\widehat{\sigma}^{2,h}_{c,n_2}}{n_{2,h,\kappa}}}} \xrightarrow{\mathcal{L}} N(0,1) \tag{6.28}$$

under the null, where $n_{1,\kappa} = \lfloor n_1^{2\kappa} \rfloor$ with $\widehat{\mu}_{n_{1,\kappa}} = n_{1,\kappa}^{-1} \sum_{(X_i,Y_i)\in\mathcal{S}^*_{n_{1,\kappa}}} \widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_{n_1})$, and $\mathcal{S}^*_{n_{1,\kappa}}$ is a random subsample of size $n_{1,\kappa}$ taken from $\mathcal{S}_{n_1}$ (see Kneip et al., 2015b for details). For the conditional part, we have similarly and as described in the preceding section, $n_{2,h,\kappa} = \lfloor n_{2,h}^{2\kappa} \rfloor$, with $\widehat{\mu}_{c,n_{2,h,\kappa}} = n_{2,h,\kappa}^{-1} \sum_{(X_i,Y_i,Z_i)\in\mathcal{S}^*_{n_{2,h,\kappa}}} \widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_{n_2})$ where $\mathcal{S}^*_{n_{2,h,\kappa}}$ is a random subsample of size $n_{2,h,\kappa}$ from $\mathcal{S}_{n_2}$.

Given a random sample $\mathcal{S}_n$, one can compute values $\widehat{\tau}_{7,n}$ or $\widehat{\tau}_{8,n}$ depending on the value of $(p+q)$. The null should be rejected whenever $1-\Phi(\widehat{\tau}_{7,n})$ or $1-\Phi(\widehat{\tau}_{8,n})$ is less than the desired test size, e.g., .1, .05, or .01, where $\Phi(\cdot)$ denotes the standard normal distribution function.

6.5 Computational Considerations

Each of the tests described in Sections 6.1, 6.2 and 6.4 require randomly splitting the original sample of size n into two parts in order to compute sample means that are independent of each other. In addition, these tests as well as the test in Section 6.3 require splitting the two parts randomly in order to obtain bias estimates. At several points in the preceding discussion, observations are assumed to be randomly ordered. In practice, however, data are often not randomly ordered when they are first obtained. Data may be sorted by firms' size or by some other criteria, perhaps one not represented by a variable or variables included in the researcher's data.

Observations can be randomly ordered by applying the modified Fisher-Yates shuffle (Fisher and Yates, 1948) described by Durstenfeld (1964). The generalized jackknife estimates of bias involve randomly splitting groups of observations many times and then averaging, and the random splitting can be accomplished by shuffling the observations before each split. One should expect little difference in the final bias estimates between two researchers who initialize their random number generators with different seeds. However, the tests of convexity, CRS, and separability require an initial split of a single sample into two parts. Conceivably, the tests may provide different results depending on how the sample is split, making if difficult to replicate results. To avoid this problem, Daraio et al. (2018) provide an algorithm for splitting the initial sample in a random but repeatable way. The algorithm ensures that Two researchers working independently with the same data will obtain the same results, even if the two researchers receive the data with different orderings of the observations. See Daraio et al. (2018) for additional details.

7 Dimension Reduction

The discussion in Sections 3.1 an 3.2 makes clear that under appropriate regularity conditions the FDH, VRS-DEA, and CRS-DEA estimators converge at rate n^{κ} where $\kappa = 1/(p+q)$, $2/(p+q+1)$ and $2/(p+q)$ (respectively). In all three cases, the convergence rates become slower—and hence estimation error increases for a fixed sample size n—as the dimensionality $(p+q)$ increases. This inverse relationship between dimensionality and convergence rates of estimators is well-known in nonparametric statistics and econometrics, and is often called the "curse of dimensionality" after Bellman (1957). In the case of FDH and DEA estimators, increasing dimensionality $(p+q)$ also affects bias, as seen in Section 5. It is perhaps less-well appreciated, but nonetheless true that dimensionality also affects the variance of partial frontier estimators such as the order-m estimators and the order-α estimators discussed in Sections 3.3 and 3.4.

Holding sample size n constant, increasing the number of inputs or the number of outputs necessarily results in a greater proportion of the sample observations lying on FDH or DEA estimates of the frontier, i.e., more observations with efficiency estimates equal to one. Applied researchers using FDH or DEA estimators have long been aware of this phenomenon. A number of ad hoc "rules of thumb" for lower bounds on the sample size n in problems with p inputs and q outputs are proposed in the literature to address the problem. For example, Bowlin (1987), Golany and Roll (1989), Vassiloglou and Giokas (1990) and Homburg (2001) propose $n \geq 2(p+q)$. Banker et al. (1989), Bowlin (1998), Friedman and Sinuany-Stern (1998) and Raab and Lichty (2002) suggest $n \geq 3(p+q)$. Boussofiane et al. (1991) offer $n \geq pq$, and Dyson et al. (2001) recommend $n \geq 2pq$. Cooper et al. (2000), Cooper et al. (2004) and Zervopoulos et al. (2012) advise $n \geq \max(pq,\ 3(p+q))$. No theoretical justification is given for any of these rules. Wilson (2018) provides evidence that the sample sizes suggested by these rules are too small to provide meaningful estimates of technical efficiency.

Wilson (2018) provides three diagnostics that can be used to warn researchers of situations where the number of inputs and outputs is excessive for a given sample size. The first diagnostic, the effective parametric sample size, is based on evaluating the number of observations m required in a parametric estimation problem (with the usual parametric convergence rate, $m^{1/2}$) to obtain estimation error of the same order that one would obtain with

n observations in a nonparametric problem with convergence rate n^{κ}. Recall that for FDH, VRS-DEA or CRS-DEA estimators, $\kappa = 1/(p+q)$, $2/(p+q+1)$ or $2/(p+q)$. To illustrate, note that Charnes et al. (1981) use the CRS-DEA estimator to examine $n = 70$ schools that use $p = 5$ inputs to produce $q = 3$ outputs. Simple calculations indicate that Charnes et al. achieve estimation error of the same order that one would attain with only 8 observations in a parametric problem. Using the VRS-DEA or FDH estimator with the Charnes et al. (1981) would result in estimation error of the same order that one would obtain in a parametric problem with 7 or 3 observations, respectively. One would likely be suspicious of estimates from a parametric problem with 8 or fewer observations, and one should be suspicious here, too. The ad hoc rules listed above suggest from 15 to 30 observations for the Charnes et al. (1981) application, which is less than half the number of observations used in their study.

With the CRS-DEA estimator, Charnes et al. (1981) obtain 19 efficiency estimates equal to 1. The VRS-DEA and FDH estimators yield 27 and 64 estimates equal to 1, respectively.[14]

Recalling that the FDH, VRS-DEA, and CRS-DEA estimators are progressively more restrictive, it is apparent that much of the inefficiency reported by Charnes et al. is due to their assumption of CRS. Under the assumptions of the statistical model (e.g., under Assumptions 2.4 and 2.5), there is not probability mass along the frontier, and hence there is zero probability of obtaining a firm with no inefficiency. The second diagnostic suggested by Wilson (2018) involves considering the number of observations that yield FDH efficiency estimates equal to 1. If this number is more than 25-50 percent of the sample size, then dimensionality may be excessive.

The third diagnostic proposed by Wilson (2018) is related to the method for reducing the dimensionality proposed by Mouchart and Simar (2002) and Daraio and Simar (2007a, pp. 148–150). As in Section 3.2, let $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote the $(p \times n)$ and $(q \times n)$ matrices of observed, input and output vectors in the sample $\mathcal{S}_n$ and suppose the rows of $\boldsymbol{X}$ and $\boldsymbol{Y}$ have been standardized by dividing each element in each row by the standard deviation of the values in each row. This affects neither the efficiency measure defined in (2.3) nor its DEA and FDH estimators, which are invariant to units of measurement. Now consider the $(p \times p)$ and $(q \times q)$ moment matrices $\boldsymbol{X}\boldsymbol{X}'$ and $\boldsymbol{Y}\boldsymbol{Y}'$. The moment matrices are by construction

[14] Here, efficiency estimates are obtained using the full sample of 70 observations. Charnes et al. split their sample into two groups according to whether schools participated in a social experiment known as "Program Follow Through." These data are analyzed further below in Section 8.

square, symmetric, and positive definite. Let $\lambda_{x1}, \ldots, \lambda_{xp}$ denote the eigenvalues of $\boldsymbol{X}\boldsymbol{X}'$ arranged in decreasing order, and let Λ_x denote the $(p \times p)$ matrix whose jth column contains the eigenvector corresponding to λ_{xj}. Define

$$R_x := \frac{\lambda_{x1}}{\sum_{j=1}^{p} \lambda_{xj}}, \tag{7.1}$$

and use the eigenvalues of $\boldsymbol{Y}\boldsymbol{Y}'$ to similarly define R_y. Let Λ_y denote the $(q \times q)$ matrix whose jth column contains the eigenvector corresponding to λ_{yj}, the jth largest eigenvalue of $\boldsymbol{Y}\boldsymbol{Y}'$.

It is well known that R_x and R_y provide measures of how close the corresponding moment matrices are to rank-one, regardless of the joint distributions of inputs and outputs. For example, if R_x=0.9, then the first principal component $\boldsymbol{X}\Lambda_{x1}$ (where Λ_{x1} is the first column of Λ_x) contains 90 percent of the independent linear information contained in the p columns of $\boldsymbol{X}$. One might reasonably replace the p inputs with this principal component, thereby reducing dimensionality from $(p + q)$ to $(1 + q)$. If, in addition, $R_y = 0.9$, then the first principal component $\boldsymbol{Y}\Lambda_{y1}$ (where Λ_{y1} is the first column of Λ_y) contains 90 percent of the independent linear information contained in the q columns of $\boldsymbol{Y}$. As with inputs, one might reasonably replace the q outputs with this principle component, further reducing dimensionality of the problem to 2. Simulation results provided by Wilson (2018) suggest that in many cases, expected estimation error will be reduced when dimensionality is reduced from $(p+q)$ to 2. Färe and Lovell (1988) show that any aggregation of inputs our outputs will distort the true radial efficiency if the technology is not homothetic, but in experiments #1 and #2 of Wilson (2018), the technology is not homothetic but any distortion is outweighed by the reduction in estimation error resulting from dimension reduction in many cases. See Wilson (2018) for specific guidelines.

Adler and Golany (2001, 2007) propose an alternative method for reducing dimensionality. In their approach, correlation matrices of *either* inputs or outputs (but not both) are decomposed using eigensystem techniques. When estimating efficiency in the output orientation, the dimensionality of the inputs can be reduced, or when estimating efficiency in the input direction, the dimensionality of the outputs can be reduced. The Adler and Golany method cannot be used when the hyperbolic efficiency measure is estimated, and can only be used with directional efficiency in special cases. The approach of Wilson (2018) can be used

in all cases. Moreover, it is well known, and is confirmed by the simulation results provided by Wilson (2018), that correlation is not a particularly meaningful measure of association when data are not multivariate normally distributed, as is often the case with production data. For purposes of estimating *radial* efficiency efficiency measures such as those defined in (2.3), (2.5) and (2.7), the issues is not whether the data are linearly related or how they are dispersed around a central point (as measured by central moments and correlation coefficients based on central moments), but rather how similar are the rays from the origin to each datum in $\mathbb{R}_+^{p+q}$ (as measured by the raw moments used by Wilson, 2018).

8 An Empirical Illustration

In this section we illustrate some of the methods described in previous sections using the "Program Follow Through" data examined by Charnes et al. (1981). All of the computations described in this section are made using the *R* programming language and the FEAR software library described by Wilson (2008).

Charnes et al. (1981) give (in their Tables 1–4) observations on $p = 5$ inputs and $q = 3$ outputs for $n = 70$ schools. The first $n_1 = 49$ of these schools participated in the social experiment known as Program Follow Through, while the remaining $n_2 = 21$ schools did not. In their Table 5, Charnes et al. report input-oriented CRS-DEA estimates (rounded to two decimal places) for the two groups of schools obtained by separating the schools into two groups depending on their participation in Program Follow Through and estimating efficiency for each group independent of the other group.

The implicit assumption of CRS by Charnes et al. (1981) is a strong assumption. Table 1 shows CRS-DEA, VRS-DEA, and FDH estimates of input-oriented efficiency for the Charnes et al. (1981) data. The observation numbers in the first column correspond to those listed by Charnes et al. in their tables; schools represented by observations numbers 1–49 participated in Program Follow Through, while the schools corresponding to observation numbers 50–70 did not. The estimates shown in Table 1 reveal that the results are sensitive to what is assumed. In particular, the CRS-DEA estimates are quite different from the VRS-DEA estimates, which in turn are quite different from the FDH estimates. From the earlier discussion in Sections 3.1 and 3.2, it is clear that both the VRS-DEA and FDH estimators remain consistent under constant returns to scale. It is equally clear that among the three

estimators, the CRS-DEA estimator is the most restrictive. The results shown in Table 1 cast some doubt on the assumption of constant returns to scale and perhaps also the assumption of a convex production set.

In addition to the sensitivity of the results in Table 1 with respect to whether the CRS-DEA, VRS-DEA or FDH estimator is used, the results reveal a more immediate problem. Among the CRS-DEA estimates shown in Table 1, 25—more than one third of the sample—are equal to one, while 33 of the VRS-DEA estimates equal one and 65 of the FDH estimates are equal to one. As discussed above in Section 7 and by Wilson (2018) this is indicative of too many dimensions for the given number of observations. Moreover, even with the assumption of constant returns to scale, the convergence rate of the CRS-DEA estimator used by Charnes et al. (1981) is $n^{1/4}$. As discussed in Section 7 and Wilson (2018), this results in an effective parametric sample size of 8. Moreover, an eigensystem analysis on the full data yields $R_x = 94.923$ and $R_y = 99.042$. The discussion in Section 7 and in Wilson (2018) makes clear the need for dimension reduction. The simulation results obtained by Wilson (2018) suggest that mean-square error is likely reduced when the 8 outputs are combined into a single principal component and the 3 outputs are combined into a single principal component using eigenvectors of the moment matrices of the inputs and outputs as described by Wilson (2018).

Table 1 also shows (input-oriented) order-α efficiency estimates (with $\alpha = 0.95$) and order-m efficiency estimates (with $m = 115$, chosen to give the number of observations *above* the order-m partial frontier similar to the number of observations lying above the order-α frontier). Note that only 3 observations lie *below* the estimated order-α frontier, and only 4 observations lie below the estimated order-m frontier (indicated by estimates less than 1). This provides further evidence that the number of dimensions is too large for the available amount of data.

Table 2 shows the eigenvectors corresponding to the largest eigenvalues for the moment matrices of inputs and outputs as discussed in Section 7. The first column of 2 gives the eigenvectors as well as values of R_x and R_y computed from the full sample with 70 observations. The second column gives similar information computed from observations 1–49 corresponding to the Program Follow Through participants, while the third column gives similar information computed from observations 50–70 for schools that did not participate

Table 1: Efficiency Estimates for Charnes et al. (1081) Data

Obs.	CRS-DEA	VRS-DEA	FDH	Order-α	Order-m
1	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.9017	0.9121	1.0000	1.0000	1.0000
3	0.9883	1.0000	1.0000	1.0000	1.0000
4	0.9024	0.9035	0.9511	1.0000	0.9523
5	1.0000	1.0000	1.0000	1.7488	1.0283
6	0.9069	0.9456	1.0000	1.2500	1.0125
7	0.8924	0.8929	1.0000	1.1399	1.0018
8	0.9148	0.9192	1.0000	1.0000	1.0000
9	0.8711	0.8877	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000
11	0.9819	1.0000	1.0000	1.0000	1.0000
12	0.9744	1.0000	1.0000	1.0000	1.0000
13	0.8600	0.8630	0.9866	0.9866	0.9866
14	0.9840	1.0000	1.0000	1.9429	1.0449
15	1.0000	1.0000	1.0000	3.8089	1.0343
16	0.9503	0.9507	1.0000	1.0000	1.0000
17	1.0000	1.0000	1.0000	1.7107	1.0126
18	1.0000	1.0000	1.0000	1.0000	1.0000
19	0.9501	0.9577	1.0000	1.0000	1.0000
20	1.0000	1.0000	1.0000	1.0000	1.0000
21	1.0000	1.0000	1.0000	1.0000	1.0000
22	1.0000	1.0000	1.0000	1.2081	1.0009
23	0.9630	0.9771	1.0000	1.0000	1.0000
24	1.0000	1.0000	1.0000	1.4075	1.0000
25	0.9764	0.9864	1.0000	1.1242	1.0006
26	0.9371	0.9425	1.0000	1.0000	1.0000
27	1.0000	1.0000	1.0000	1.0000	1.0000
28	0.9443	0.9903	1.0000	1.3333	1.0025
29	0.8417	0.9325	1.0000	1.3748	1.0155
30	0.9025	0.9119	1.0000	1.3093	1.0059
31	0.8392	0.8520	0.9915	1.0000	0.9918
32	0.9070	1.0000	1.0000	1.7778	1.0363
33	0.9402	0.9578	1.0000	1.0000	1.0000
34	0.8521	0.8645	1.0000	1.0000	1.0000
35	1.0000	1.0000	1.0000	1.0000	1.0000

Table 1: Efficiency Estimates for Charnes et al. (1081) Data (continued)

Obs.	CRS-DEA	VRS-DEA	FDH	Order-α	Order-m
36	0.8032	0.8033	1.0000	1.0064	1.0007
37	0.8614	0.8692	1.0000	1.0000	1.0000
38	0.9485	1.0000	1.0000	2.0205	1.0363
39	0.9352	0.9438	1.0000	1.0000	1.0000
40	1.0000	1.0000	1.0000	1.4633	1.0016
41	0.9468	0.9526	1.0000	1.1545	1.0000
42	0.9474	0.9531	1.0000	1.3333	1.0025
43	0.8708	0.8752	1.0000	1.0000	1.0000
44	1.0000	1.0000	1.0000	1.0000	1.0000
45	0.8916	1.0000	1.0000	2.0000	1.0471
46	0.9087	0.9283	1.0000	1.0000	1.0000
47	1.0000	1.0000	1.0000	1.0000	1.0000
48	1.0000	1.0000	1.0000	2.2037	1.0305
49	1.0000	1.0000	1.0000	2.2885	1.0172
50	0.9575	0.9587	1.0000	1.0000	1.0000
51	0.9205	0.9277	1.0000	1.0000	1.0000
52	1.0000	1.0000	1.0000	1.0000	1.0000
53	0.8768	0.8970	0.9872	0.9872	0.9872
54	1.0000	1.0000	1.0000	1.0000	1.0000
55	1.0000	1.0000	1.0000	1.0000	1.0000
56	1.0000	1.0000	1.0000	1.0000	1.0000
57	0.9260	0.9269	1.0000	1.0000	1.0000
58	1.0000	1.0000	1.0000	1.0000	1.0000
59	0.9223	1.0000	1.0000	1.0000	1.0000
60	0.9815	0.9981	1.0000	1.0000	1.0000
61	0.8818	0.9012	1.0000	1.0000	1.0000
62	1.0000	1.0000	1.0000	1.0000	1.0030
63	0.9611	0.9634	1.0000	1.0000	1.0000
64	0.9168	0.9373	1.0000	1.0000	1.0000
65	0.9775	0.9775	1.0000	1.0000	1.0000
66	0.9259	0.9412	0.9761	0.9761	0.9761
67	0.9271	0.9492	1.0000	1.0000	1.0000
68	1.0000	1.0000	1.0000	1.0000	1.0000
69	1.0000	1.0000	1.0000	1.0000	1.0000
70	0.9475	0.9640	1.0000	1.0000	1.0000

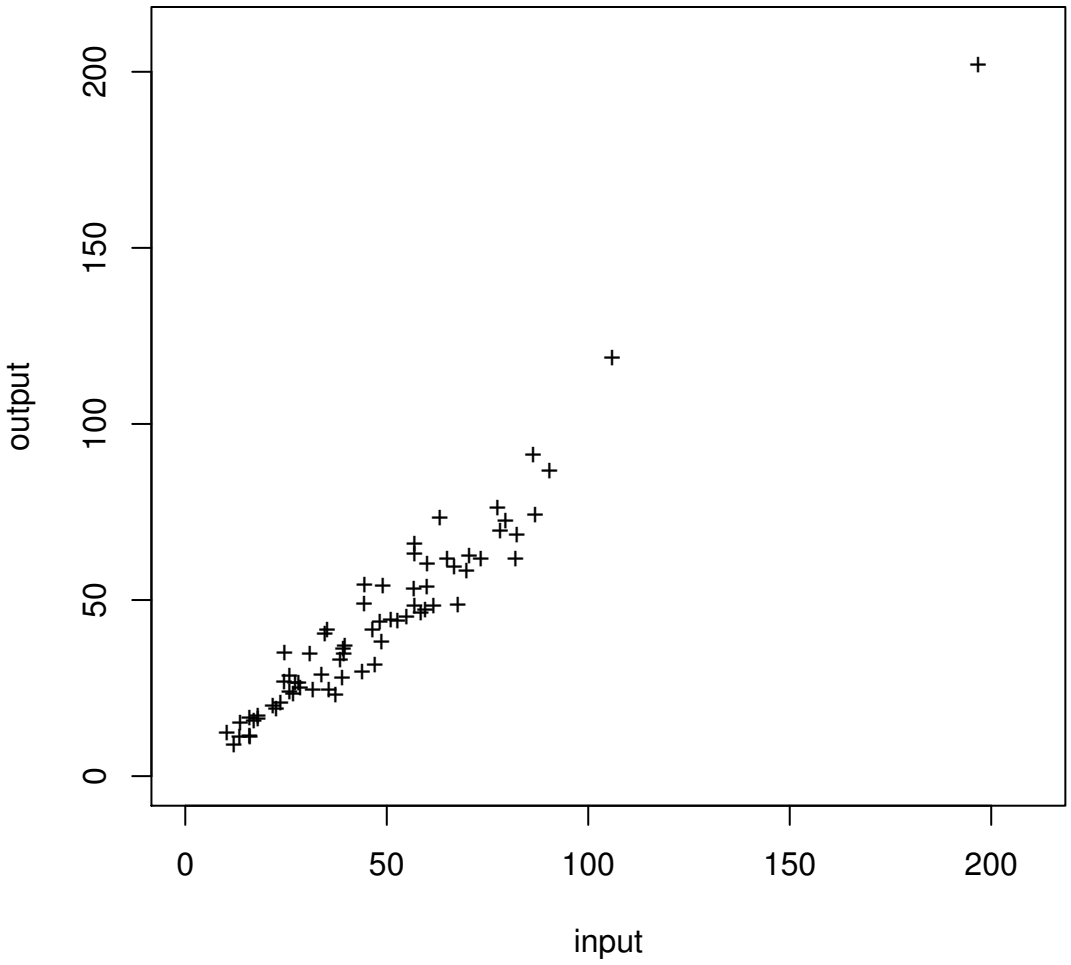
in Program Follow Through. The results are similar across the three columns of Table 2, and so it appears reasonable to use the eigenvectors computed from the full sample to compute principal components. Doing so for both inputs and outputs reduces dimensionality from 8 to 2.

Table 2: Eigenvectors of Input and Output Moment Matrices

	(1)	(2)	(3)
Λ_{x1}	0.3829	0.3855	0.3799
	0.4662	0.4538	0.4493
	0.4700	0.4479	0.4541
	0.4661	0.4635	0.4394
	0.4450	0.4796	0.5046
R_x	94.9235	96.6599	93.4648
Λ_{y1}	0.5602	0.5702	0.5647
	0.5593	0.5771	0.5548
	0.6110	0.5846	0.6110
R_y	99.0422	99.1462	99.3443

The first principal components of inputs and outputs based on moment matrices as described above are plotted in Figure 1. The plot in Figure 1 reveals two dissimilar observations. The observation lying in the upper right corner of the plot is observation number 59, and the observation lying at (105.93, 118.93) is observation number 44. Both of these observations are flagged by Wilson (1993, 1995) and Simar (2003) as outliers. Note that the transformation from the original $(p + q)$-dimensional space of inputs and outputs to the 2-dimensional space of first principal components amounts to an affine function. It is well-known that the image of a convex set under an affine function is also convex, and that the affine transformation preserves extreme points (e.g., see Boyd and Vandenberghe, 2004). Therefore the method for dimension reduction discussed in Section 7 and Wilson (2018) has an additional benefit. Reducing dimensionality to only 2 dimensions permits simple scatter plots of the transformed data, which can reveal outliers in the data that are more difficult to detect in higher dimensions.

Figure 1: Charnes et al. (1981) Data after Principal Components Transformation



Applying the CRS-DEA, VRS-DEA, FDH, order-α and order-m estimators to the transformed data yields the results shown in Tables 3–4. The estimates in Table 3 are obtained by applying the estimators separately on the two groups of observations (i.e., those participating in Program Follow Through, corresponding to observations 1–49, and those not participating, corresponding to observations 50–70). The estimates in Table 4 are obtained from the full sample, ignoring program participation. In both Tables 3–4 the partial efficiency estimates are computed with $\alpha = 0.95$ and $m = 115$ as in Table 1.

As an illustration, consider the results for observation no. 5 in Table 3. The CRS-DEA and VRS-DEA estimates suggest that this school could produce the same level of outputs using only 69.91 percent or 98.65 percent of its observed levels of inputs, respectively. The FDH estimate indicates that school no. 5 is on the estimated frontier, and by itself does not suggest that input levels could be reduced without reducing output levels. At the same time, however, the FDH estimate, as well as the CRS-DEA and VRS-DEA estimates, are biased upward toward 1. The order-m estimate, by contrast, suggests that school no. 5 would have to increase its input usage by 1.06 percent in order to meet the estimated *expected* input usage among 115 randomly chosen schools producing at least as much output as school no. 5. Similarly, the order-α estimate suggests that school no. 5 would have to increase its input usage by 26.70 percent while holding output levels fixed in order to have a probability of 0.05 of being dominated by a school producing more output while using less input that school no. 5. In the input orientation, the partial efficiency estimators are necessarily weakly greater than the FDH estimates. After reducing dimensionality to only two dimensions, the partial efficiency estimators have the same convergence rate (i.e., $n^{1/2}$) as the FDH estimators, but the partial efficiency estimates remain less sensitive to outliers than the FDH estimates.

Comparing results across the five columns in either Table 3 or Table 4 reveals substantial differences between estimates obtained with different estimators. This raises the question of whether constant returns to scale (assumed by Charnes et al.) is an appropriate assumption supported by the data, or whether even convexity of the production set is an appropriate assumption. In addition, comparing estimates in Table 3 with corresponding estimates in Table 4 reveals some large differences, but not in all cases. This raises the question of whether the two groups of schools have the same mean efficiency or whether they face the same frontier. Charnes et al. allow for different frontiers, but if the schools face the same

frontier, using the full sample for estimation would reduce estimation error. Rather than making arbitrary assumptions, the statistical results described in earlier sections can be used to let the data speak to what is appropriate.

In order to address the question of whether the two groups of schools face the same frontier, we apply the test of separability developed by Daraio et al. (2018) using the reduced-dimensional, transformed data. Note that here, the "environmental" variable is binary; i.e., schools either participate in Program Follow Through or they do not. Hence no bandwidths are needed; instead, we follow the method outlined in Daraio et al. (2018, separate Appendix C) and using the difference-in-means test described by Kneip et al. (2016). To employ the test, we first randomly sort the observations using the randomization algorithm appearing in Daraio et al. (2018, separate Appendix D). We then divide the observations into two sub-samples, taking the first 35 randomly sorted observations as group 1, and the remaining observations as group 2. Using the observations in group 1, we estimate efficiency using the input-oriented FDH estimator and compute the sample mean (0.8948) and sample variance (0.01424) of these estimates. We then estimate the bias (0.2540) of this sample mean using the generalized jackknife method described by Daraio et al. (2018).

For the 35 observations in group 2, we have 23 observations on schools participating in Program Follow Through (group 2a), and 12 observations not participating (group 2b). We again use the FDH estimator to estimate input-oriented efficiency, but we apply the estimator as well as the generalized jackknife independently on the two sub-groups (2a and 2b). We then compute the sample mean (across 35 observations in group 2) of the efficiency estimates (0.9170), the sample variance (0.01408) and the bias correction (0.2011) as described by Daraio et al. (2018, separate Appendix C). This yields a value of 2.6278 for the test statistic given by the input-oriented analog of (6.19), and hence the p-value is 0.004298.[15] Hence the null hypothesis of separability is soundly rejected by the data. The data provide evidence suggesting that Program Follow Through schools face a different frontier than the non-Program Follow Through schools.

We use the FDH estimator to test separability since the FDH estimator does not require an assumption of convexity, which has not yet been tested. Given that separability is rejected,

[15] The test is a one-sided test, since by construction the mean of the input-oriented efficiency estimates for group 1 is less than the similar mean for group 2 under departures from the null hypothesis of separability.

Table 3: Efficiency Estimates from Transformed Data, Split Sample

Obs.	CRS-DEA	VRS-DEA	FDH	Order-α	Order-m
1	0.7830	0.8332	1.0000	1.0000	1.0000
2	0.6975	0.7037	0.7804	0.7821	0.7804
3	0.8212	0.8351	0.9466	0.9466	0.9466
4	0.5882	0.6285	0.6642	0.7934	0.6705
5	0.6991	0.9865	1.0000	1.2670	1.0106
6	0.7664	0.8357	0.9485	1.0000	0.9502
7	0.5560	0.5892	0.7035	0.8026	0.7075
8	0.5876	0.5927	0.6564	0.6578	0.6564
9	0.6478	0.6550	0.7449	0.7466	0.7450
10	0.8031	0.8425	1.0000	1.0000	1.0000
11	0.7765	0.7923	0.8753	0.8753	0.8753
12	0.9082	0.9300	0.9988	0.9988	0.9988
13	0.6426	0.6483	0.7203	0.7219	0.7203
14	0.7434	0.8748	0.8869	1.0000	0.8939
15	0.8987	0.9652	1.0000	1.0540	1.0013
16	0.6882	0.7022	0.7753	0.7753	0.7753
17	0.9038	0.9631	1.0000	1.1945	1.0066
18	0.9048	0.9119	1.0000	1.0000	1.0000
19	0.7281	0.7556	0.9912	0.9912	0.9912
20	1.0000	1.0000	1.0000	1.0000	1.0000
21	0.9499	0.9788	1.0000	1.0000	1.0000
22	0.9248	0.9622	1.0000	1.1408	1.0037
23	0.7470	0.7792	0.9752	0.9752	0.9752
24	0.9657	0.9872	1.0000	1.2599	1.0041
25	0.7320	0.7483	0.9544	0.9565	0.9544
26	0.7275	0.7379	0.8523	0.8523	0.8523
27	0.9016	0.9020	0.9070	0.9070	0.9070
28	0.7217	0.7513	0.7846	0.8950	0.7860
29	0.5767	0.8218	0.8395	1.0000	0.8469
30	0.6342	0.6902	0.7749	0.8167	0.7775
31	0.5660	0.6156	0.6884	0.7255	0.6899
32	0.6149	1.0000	1.0000	1.3282	1.0120
33	0.6992	0.7314	0.8925	0.8925	0.8925
34	0.6849	0.6931	0.8146	0.8146	0.8146
35	0.6809	0.7052	0.9412	0.9412	0.9412

Table 3: Efficiency Estimates from Transformed Data, Split Sample (continued)

Obs.	CRS-DEA	VRS-DEA	FDH	Order-α	Order-m
36	0.5122	0.5617	0.6583	0.6940	0.6606
37	0.7000	0.7447	0.9137	1.0000	0.9153
38	0.8601	1.0000	1.0000	1.5371	1.0115
39	0.7344	0.7350	0.7422	0.7422	0.7422
40	0.7668	0.7927	0.8894	1.0000	0.8913
41	0.6860	0.6975	0.8429	0.8449	0.8430
42	0.7554	0.7829	0.8983	1.0000	0.8992
43	0.6502	0.6582	0.7593	0.7610	0.7593
44	0.9147	1.0000	1.0000	1.0000	1.0000
45	0.7882	0.9108	1.0000	1.4389	1.0223
46	0.6171	0.6299	0.6937	0.6937	0.6937
47	0.7679	0.7691	0.7839	0.7839	0.7843
48	0.7646	0.9111	0.9365	1.0544	0.9389
49	0.7276	0.7880	0.8605	0.9070	0.8629
50	0.6235	0.7428	0.8968	0.8968	0.8968
51	0.6003	0.6483	0.9650	0.9650	0.9650
52	0.8149	1.0000	1.0000	1.0000	1.0000
53	0.6071	0.6104	0.6413	0.6413	0.6413
54	0.7412	0.9461	1.0000	1.0000	1.0000
55	0.8248	0.8744	1.0000	1.0000	1.0000
56	0.6846	0.7058	0.9021	0.9021	0.9021
57	0.6398	0.6962	1.0000	1.0000	1.0000
58	1.0000	1.0000	1.0000	1.0000	1.0000
59	0.7205	1.0000	1.0000	1.0000	1.0000
60	0.6700	0.6901	0.8763	0.8763	0.8763
61	0.5146	0.6481	0.6481	0.8531	0.6489
62	0.7948	0.8934	1.0000	1.0000	1.0019
63	0.6240	0.6645	1.0000	1.0000	1.0000
64	0.5508	0.5703	0.7119	0.7119	0.7119
65	0.6509	0.6979	1.0000	1.0000	1.0000
66	0.4744	0.4791	0.5232	0.5232	0.5232
67	0.5809	0.6384	1.0000	1.0000	1.0000
68	0.6137	0.6706	1.0000	1.0000	1.0000
69	0.8511	1.0000	1.0000	1.0000	1.0000
70	0.6157	0.6453	0.9193	0.9193	0.9193

Table 4: Efficiency Estimates from Transformed Data, Full Sample

Obs.	CRS-DEA	VRS-DEA	FDH	Order-α	Order-m
1	0.6733	0.8277	0.9555	0.9555	0.9555
2	0.5997	0.6640	0.7804	0.7821	0.7808
3	0.7061	0.8202	0.9466	0.9466	0.9466
4	0.5057	0.5181	0.6328	0.7934	0.6368
5	0.6011	0.7695	0.7695	1.1865	0.8049
6	0.6590	0.6876	0.9485	0.9997	0.9496
7	0.4780	0.4861	0.5611	0.7893	0.5697
8	0.5053	0.5596	0.6564	0.6578	0.6564
9	0.5570	0.6124	0.7449	0.7466	0.7449
10	0.6905	0.8341	1.0000	1.0000	1.0000
11	0.6677	0.7788	0.8753	0.8753	0.8753
12	0.7809	0.9151	0.9988	0.9988	0.9988
13	0.5525	0.6114	0.7203	0.7219	0.7203
14	0.6392	0.7102	0.8869	1.2083	0.9043
15	0.7728	0.7953	1.0000	1.0540	1.0016
16	0.5917	0.6903	0.7753	0.7753	0.7753
17	0.7771	0.7942	0.9527	1.1945	0.9561
18	0.7780	0.8639	1.0000	1.0023	1.0000
19	0.6261	0.7459	0.8079	0.8079	0.8079
20	0.8598	0.9778	1.0000	1.0000	1.0000
21	0.8168	0.9646	1.0000	1.0000	1.0000
22	0.7951	0.7954	0.7975	1.1219	0.8057
23	0.6423	0.7704	0.7949	0.7949	0.7949
24	0.8303	0.8808	1.0000	1.2599	1.0106
25	0.6294	0.6679	0.9544	0.9565	0.9547
26	0.6255	0.7243	0.8523	0.8523	0.8523
27	0.7753	0.8805	0.9070	1.0000	0.9079
28	0.6205	0.6210	0.6257	0.8802	0.6311
29	0.4958	0.6460	0.6460	0.9961	0.6734
30	0.5453	0.5680	0.7749	0.8167	0.7755
31	0.4867	0.5066	0.6884	0.7255	0.6894
32	0.5287	0.8614	0.8614	1.1338	0.8823
33	0.6012	0.7235	0.8925	0.8925	0.8925
34	0.5889	0.6799	0.8146	0.8146	0.8146
35	0.5855	0.6958	0.7672	0.7672	0.7672

Table 4: Efficiency Estimates from Transformed Data, Full Sample (continued)

Obs.	CRS-DEA	VRS-DEA	FDH	Order-α	Order-m
36	0.4404	0.4619	0.6583	0.6938	0.6589
37	0.6019	0.6142	0.7287	1.0000	0.7362
38	0.7395	0.8178	1.0000	1.4118	1.0334
39	0.6314	0.7164	0.7422	0.8183	0.7422
40	0.6593	0.6738	0.8746	0.8894	0.8761
41	0.5898	0.6368	0.8429	0.8449	0.8430
42	0.6495	0.6578	0.8834	0.8983	0.8851
43	0.5591	0.6121	0.7593	0.7610	0.7593
44	0.7865	1.0000	1.0000	1.0000	1.0000
45	0.6778	0.7452	1.0000	1.3175	1.0344
46	0.5306	0.6193	0.6937	0.6937	0.6937
47	0.6603	0.7475	0.7839	0.8642	0.7850
48	0.6574	0.7341	0.9365	1.0559	0.9469
49	0.6256	0.6488	0.8605	0.9070	0.8615
50	0.6235	0.7291	0.8073	0.8073	0.8073
51	0.6003	0.6483	0.9650	1.0507	0.9709
52	0.8149	0.9787	1.0000	1.0000	1.0000
53	0.6071	0.6104	0.6413	0.9021	0.6485
54	0.7412	0.9169	1.0000	1.0000	1.0000
55	0.8248	0.8687	1.0000	1.0168	1.0008
56	0.6846	0.7058	0.8984	0.9469	0.8993
57	0.6398	0.6897	0.9197	0.9217	0.9205
58	1.0000	1.0000	1.0000	1.4067	1.0133
59	0.7205	1.0000	1.0000	1.0000	1.0000
60	0.6700	0.6901	0.8728	0.9199	0.8736
61	0.5146	0.6481	0.6481	1.0000	0.6789
62	0.7948	0.8934	1.0000	1.3189	1.0327
63	0.6240	0.6645	1.0000	1.0404	1.0050
64	0.5508	0.5681	0.7119	0.7239	0.7125
65	0.6509	0.6979	1.0000	1.1282	1.0113
66	0.4744	0.4791	0.5232	0.7359	0.5296
67	0.5809	0.6318	0.8088	0.8107	0.8089
68	0.6137	0.6640	0.8706	0.8726	0.8706
69	0.8511	1.0000	1.0000	1.6466	1.0481
70	0.6157	0.6453	0.9155	0.9650	0.9164

we next apply the convexity test of Kneip et al. (2016) separately and independently on the 49 participating schools and the 21 non-participating schools, again using the two-dimensional, transformed data. For the participating and non-participating schools we obtain values 3.4667 and -0.6576 (respectively) of the test statistic given in (6.7). The corresponding p-values are 0.0003 and 0.7446, and hence we firmly reject convexity of the production set for participating schools, but we fail to reject convexity for the non-participating schools.

Next, for the non-participating schools, we test constant returns to scale against the alternative hypothesis of variable returns to scale using the method of Kneip et al. (2016). Since convexity is rejected for the Program Follow Through schools, there is no reason to test constant returns to scale. As with the other tests, we also work here again with the two-dimensional, transformed data. We obtain the value -1.7399 for the input-oriented analog of the test statistic in (6.14), with a corresponding p value of 0.9591. Hence we do not reject constant returns to scale for the non-participating schools.

Taken together, the results of the tests here suggest that efficiency should be estimated separately and independently for the group of participating and the group of non-participating schools. In addition, the results indicate that the FDH estimator should be used to estimate efficiency for the participating schools. On the other hand, we do not find evidence to suggest that the CRS-DEA estimator is inappropriate for the non-participating schools.

9 Nonparametric Panel Data Frontier

9.1 Basic Ideas: Conditioning on Time

One possible approach, developed in Mastromarco and Simar (2015) in a macroeconomic setup, is to extend the basic ideas of Cazals et al. (2002) and Daraio and Simar (2005), described in Section 4, to a dynamic framework to allow introduction of a time dimension.

Consider a generic input vector $X \in \mathbb{R}^p_+$, a generic output vector $Y \in \mathbb{R}^q_+$ and denote by $Z \in \mathbb{R}^d$ the generic vector of environmental variables. Mastromarco and Simar (2015) consider the time T as a conditioning variable and define for each time period t, Ψ_t, the attainable set at time t as the set of combinations of inputs and outputs feasible at time t. $\Psi_t \subset \mathbb{R}^{p+q}_+$ is the support of (X, Y) at time t, whose distribution is completely determined

by

$$H^t_{X,Y}(x,y) = \text{Prob}\left(X \leq x, Y \geq y \,|\, T = t\right), \tag{9.1}$$

which is the probability of being dominated for a production plan (x, y) at time t. Finally, when considering the presence of additional external factors Z, the attainable set is defined as $\Psi^z_t \subseteq \Psi_t \subset \mathbb{R}^{p+q}_+$ defined as the support of the conditional probability

$$H^t_{X,Y|Z}(x,y|z) = \text{Prob}\left(X \leq x, Y \geq y \,|\, Z = z, T = t\right). \tag{9.2}$$

In this framework, assuming free disposability of inputs and outputs, the conditional output oriented Farrell-Debreu technical efficiency of a production plan $(x, y) \in \Psi^z_t$, at time t facing conditions z, is defined in (4.5)–(4.7) as

$$\begin{aligned} \lambda_t\left(x,y|z\right) &= \sup\left\{\lambda | (x,\lambda y) \in \Psi^z_t\right\} = \sup\left\{\lambda | H^t_{Y,X|Z}\left(x,\lambda y|z\right) > 0\right\} \\ &= \sup\{\lambda | S^t_{Y|X,Z}(\lambda y|x,z) > 0\}. \end{aligned} \tag{9.3}$$

where $S^t_{Y|X,Z}(y|x,z) = \text{Prob}(Y \geq y|X \leq x, Z = z, T = t)$.[16]

Suppose we have a sample $\mathcal{X}_{n,s} = \{(X_{it}, Y_{it}, Z_{it})\}^{n,s}_{i=1,\, t=1}$ comprised of panel data for n firms observed over s periods. Then the unconditional and conditional attainable sets can be estimated. Assuming that the true attainable sets are convex and under free disposability of inputs and outputs, (similar to Daraio and Simar, 2007b), the DEA estimators at time t and facing the condition $Z = z$ is given by

$$\widehat{\Psi}^z_{\text{DEA},t} = \{(x,y) \in \mathbb{R}^p_+ \times \mathbb{R}^q_+ \mid y \leq \sum_{j \in \mathcal{J}(z,t)} \gamma_j Y_j,\ x \geq \sum_{j \in \mathcal{J}(z,t)} \gamma_j X_j, \gamma \geq 0,\ \sum_{j \in \mathcal{J}(z,t)} \gamma_j = 1\} \tag{9.4}$$

where $\mathcal{J}\left(z,t\right) = \{j = (i,v) \mid z - h_z < Z_{i,v} < z + h_z; t - h_t < v < t + h_t\}$ and h_z and h_t are bandwidths of appropriate size selected by data-driven methods. The set $\mathcal{J}(z,t)$ describes the localizing procedure to estimate the conditional DEA estimates, and determines the observations in a neighborhood of (z,t) that will be used to compute the local DEA estimate. Here, only the variables (t,z) require smoothing and appropriate bandwidths, see Section 4.2 above for details and Daraio et al. (2018) for discussion of practical aspects of bandwidth selection.

[16] For efficiency measures, we only focus the presentation on the output orientation, the same could be done for any other orientation (input, hyperbolic, directional distance) (see, Daraio and Simar, 2007a; Bădin et al., 2010, 2012; Wilson, 2011; Simar and Vanhems, 2012; and Simar et al., 2012).

The estimator of the output conditional efficiency measure at time t is then obtained by substituting $\widehat{\Psi}^z_{DEA,t}$ for Ψ^z_t in (9.3). In practice, an estimate is computed by solving the linear program

$$\widehat{\lambda}_{\mathrm{DEA},t}\left(x,y\mid z\right)=\sup\left\{\lambda\mid\lambda y\leq\sum_{j\in\mathcal{J}(z,t)}\gamma_jY_j,\ x\geq\sum_{j\in\mathcal{J}(z,t)}\gamma_jx_j,\ \gamma\geq0,\ \sum_{j\in\mathcal{J}(z,t)}\gamma_j=1\right\} \tag{9.5}$$

If the convexity of the sets Ψ^z_t is questionable, it is better to use the FDH approach described in Section 4.2 relying only on the free disposability of the inputs and outputs. It particularizes here as follows

$$\widehat{\Psi}^z_{\mathrm{FDH},t}=\{(x,y)\in\mathbb{R}^p_+\times\mathbb{R}^q_+\mid y\leq Y_j,\ x\geq X_j,\ j\in\mathcal{I}(z,t)\} \tag{9.6}$$

where $\mathcal{I}\left(z,t\right)=\{j=(i,v)\mid z-h_z<Z_{i,v}<z+h_z;t-h_t<v<t+h_t\cap X_{i,v}\leq x\}$. The conditional output FDH estimator turns out to be simply defined as, see (4.13),

$$\widehat{\lambda}_{\mathrm{FDH},t}(x,y\mid z)=\max_{i\in\mathcal{I}(z,t)}\left(\min_{j=1,\ldots,p}\left(\frac{Y_i^j}{y^j}\right)\right), \tag{9.7}$$

The statistical properties of these nonparametric estimators are well known as discussed in Section 4.2. To summarize, these estimators are consistent and converge to some non-degenerate limiting distributions to be approximated by bootstrap techniques. As discussed earlier, the rates of convergence become slower with increasing numbers of inputs and outputs, illustrating the curse of dimensionally. The situation is even jeopardized if the dimension of Z increases.

Having these estimators and to bring some insights on the effect of Z and T on the efficiency scores, Mastromarco and Simar (2015) analyze, in a second stage, the average behavior of $\lambda_t(X,Y\mid Z)$ at period t and conditionally on $Z=z$. The regression aims to analyze the behavior of $E\left(\widehat{\lambda}_t(X,Y\mid Z=z)\right)$, where $\widehat{\lambda}_t$ is either the DEA or the FDH estimators defined above, as a function of z and t. For this purpose they suggest to estimate the flexible nonparametric location-scale regression model

$$\widehat{\lambda}_t(X_{it},Y_{it}\mid Z_{it})=\mu(Z_{it},t)+\sigma(Z_{it},t)\varepsilon_{it} \tag{9.8}$$

where $E\big(\varepsilon_{it}|Z_{it},t\big)=0$ and $\mathrm{VAR}\big(\varepsilon_{it}|Z_{it},t\big)=1$. This model captures both the location $\mu(z,t)=E\big(\widehat{\lambda}_t(X_{it},Y_{it}|Z_{it}=z)\big)$ and the scale effect $\sigma^2(z,t)=\mathrm{VAR}\big(\widehat{\lambda}_t(X_{it},Y_{it}|Z_{it}=z)\big)$.

The nonparametric function $\mu(z,t)$ are usually estimated by local linear techniques and $\sigma^2(z,t)$ by local constant techniques on the squares of the residuals obtained when estimating $\mu(z,t)$ (see Mastromarco and Simar (2015) for details).

As a byproduct of the analysis, the scaled residual error term for a given particular production plan (X_{it}, Y_{it}, Z_{it}) at time t is computed as

$$\widehat{\varepsilon}_{it} = \frac{\lambda_t \left(X_{it}, Y_{it} \mid Z_{it}\right) - \mu \left(Z_{it}, t\right)}{\sigma \left(Z_{it}, t\right)}. \tag{9.9}$$

This unexplained part of the conditional efficiency measure (called "pure efficiency" in Bădin et al., 2012), cleanses efficiency scores from external effects (here, time T and Z). The pure efficiency measure provides then a better indicator by which to assess the economic performance of production units over time and allows the ranking of production units facing different environmental factors at different time periods.

Mastromarco and Simar (2015) apply these panel techniques in a macroeconomic setup, where they use a dataset of 44 countries (26 OECD countries and 18 developing countries) over 1970-2007. In their macroeconomic cross-country framework, where countries are producers of output (i.e., GDP) given inputs (capital, labor; see Mastromarco and Simar, 2015 for a detailed description of the data) and technology, inefficiency can be identified as the distance of the individual production from the frontier estimated by the maximum output of the reference country regarded as the empirical counterpart of an optimal boundary of the production set. Inefficiencies generally reflect a sluggish adoption of new technologies, and thus efficiency improvement will represent productivity catch-up via technology diffusion.

Mastromarco and Simar (2015) explore the channels under which FDI fosters productivity by disentangling the impact of this factor on the production process and its components: impact on the attainable production set (inputoutput space) and the impact on the distribution of efficiencies. In particular they want to clarify the effect of time and foreign direct investment (FDI) on the catching-up process which is related to productivity gains and so to productive efficiency of the countries. They review the literature in the field (both parametric and non-parametric).

The second stage regression described above cleanses efficiency scores from external effects (time and FDI), and this enables Mastromarco and Simar (2015) to eliminate the common time factor effect, as economic cycles, in a very flexible and robust way (the location-scale

model). Their pure efficiency measure provides a better indicator to assess the economic performance of production units over time and in their macroeconomic framework the "pure efficiency" represents a new measure of the Solow residual. They conclude that both time and FDI plays a role in the catching-up process, clarifying from an empirical point of view some theoretical debate on these issues.

Note that the approach developed in this section could also have been applied with robust versions of the DEA or FDH estimators by using rather the partial frontiers, the order-m and the order-α frontiers derived in Section 2.4. This is particularly useful if the cloud of data points contains extreme data points or outliers which may hide the true relationship between the variables, see e.g. the discussion in Section 5.4 in Daraio and Simar (2007a).

9.2 Cross-Section Dependence in Panel Data Frontier Analysis

When analyzing a panel of data, we might also expect some cross-sectional dependence between the units, this is particularly true for macroeconomic data but also for a panel of data on firms in a country, etc. In a macroeconomic setup, Mastromarco and Simar (2018) propose a flexible, nonparametric, two-step approach to take into account cross section dependence due to common factors attributable to global shocks to the economy. It easy to extend the approach for microeconomic data. They use again conditional measures where now they condition not only to external environmental factors (Z) but also to some estimates of a "common time factor" that is supposed to capture the correlation among the units. By conditioning on the latter, they eliminate the effect of these factors on the production process and so mitigate the problem of cross section dependence. To define this common time factor, they follow the approach Pesaran (2006) and Bai (2009), where it is shown that an unobserved common time factor, ξ_t, can be consistently proxied by cross-sectional averages of inputs and the outputs at least asymptotically, as $n, s \to \infty$, and $s/n \to K$ where K is a finite positive (≥ 0) constant. So the idea is to consider $F_t = (t, X_{\cdot t}, Y_{\cdot t})$ as a proxy for the unobserved nonlinear and complex common trending patterns.[17]

So, at this stage we have now a sample of observations $\{(X_{it}, Y_{it}, Z_{it}, F_t)\}_{i=1,t=1}^{n,s}$. For estimating the conditional measures, Mastromarco and Simar (2018) follow the approach

[17] Here, we use the standard notation where a dot in a subscript signifies an average over the corresponding index.

suggested by Florens et al. (2014) which so far has been developed for univariate Y, but the multivariate case should not create any theoretical issues. This approach avoids direct estimation of the conditional survival function $S_{Y|X,Z,F_t}(y \mid x, z, f_t)$. As observed by Florens et al., the procedure reduces the impact of the curse of dimensionality (through the conditioning variables Z, F_t) and requires smoothing in these variables in the *center* of the data cloud instead of smoothing at the frontier where data are typically sparse and estimators are more sensitive to outliers. Moreover, the inclusion of time factor $F_t = (t, X_{\cdot t}, Y_{\cdot t})$ enables elimination of the common time factor effect in a flexible, nonparametric location-scale model. The statistical properties of the resulting frontier estimators are established by Florens et al. (2014).

Assume the data are generated by the nonparametric location-scale regression model

$$\left\{ \begin{array}{ll} X_{it} & = \mu_x(Z_{it}, F_t) + \sigma_x(Z_{it}, F_t)\varepsilon_{x,it} \\ Y_{it} & = \mu_y(Z_{it}, F_t) + \sigma_y(Z_{it}, F_t)\varepsilon_{y,it} \end{array} \right. , \tag{9.10}$$

where μ_x, σ_x and ε_x each have p components and, for ease of notation, the product of vectors is component-wise. So the first equation in (9.10) represents p relations, one for each component of X. We assume that each element of ε_x and ε_y have mean zero and standard deviation equal to 1. The model also assumes that $(\varepsilon_x, \varepsilon_y)$ is independent of (Z, F_t). This model allows the capture for any (z, f_t) and for each input $j = 1, \ldots, p$ and for the output, the locations $\mu_x^{(j)}(z, f_t) = E\left(X^{(j)} \mid Z = z, F_t = f_t\right)$, $\mu_y(z, f_t) = E\left(Y \mid Z = z, F_t = f_t\right)$ and the scale effects $\sigma_x^{(j),2}(z, f_t) = \mathrm{VAR}\left(X^{(j)} | Z = z, F_t = f_t\right)$, $\sigma_y^2(z, t) = \mathrm{VAR}\left(Y | Z = z, F_t = f_t\right)$ of the environmental and common factors on the production plans. Here again, for a vector a, $a^{(j)}$ denotes its j-th component.

The production frontier can be estimated in two stages as proposed by Florens et al. (2014). In the first stage, the location functions $\mu_\ell(z_{it}, f_t)$ in (9.10) are estimated by local linear methods. Then the variance functions $\sigma_\ell^2(z_{it}, f_t)$ are estimated by regressing the squares of the residuals obtained from the first local linear regression on (z, f_t). For the variance functions, a local constant estimators is used to avoid negative values of the estimated variances.

The first-stage estimation yields the residuals

$$\widehat{\varepsilon}_{x,it} = \frac{X_{it} - \widehat{\mu}_x(Z_{it}, F_t)}{\widehat{\sigma}_x(Z_{it}, F_t)} \tag{9.11}$$

and

$$\widehat{\varepsilon}_{y,it} = \frac{Y_{it} - \widehat{\mu}_y(Z_{it}, F_t)}{\widehat{\sigma}_y(Z_{it}, F_t)}, \tag{9.12}$$

where for ease of notation, a ratio of two vectors is understood to be component-wise. These residuals amount to whitened inputs and output obtained by eliminating the influence of the external and other environmental variables as common factors. To validate the location-scale model, Florens et al. (2014) propose a bootstrap based testing procedure to test the independence between $(\widehat{\varepsilon}_{x,it}, \widehat{\varepsilon}_{y,it})$ and (Z_{it}, F_t), i.e. the independence of whitened inputs and output from the external and global effects.

Note that here, for finite sample the independence between $(\varepsilon_x, \varepsilon_y)$ and of (Z, F_t) is not verified in finite samples, as, e.g. $\mathrm{COV}(Y_{.t}, \varepsilon_{y,it}) = \sigma_y \mathrm{COV}(\varepsilon_{.t}, \varepsilon_{y,it}) = \sigma_y \frac{\mathrm{VAR}(\varepsilon_{y,it})}{n}$, but since the latter converge to zero as $n \to \infty$, this does not contradict the asymptotic independence assumed by the model.

In the second stage, a production frontier is estimated for the whitened output and inputs given by (9.11) and (9.12). Hence for each observation (i, t) a measure of "pure" efficiency is obtained. This approach accommodates both time and cross-sectional dependence and yields more reliable estimates of efficiency.[18] Moreover, as observed by Florens et al. (2014), by cleaning the dependence of external factors in the first stage, the curse of dimensionality due to the dimension of the external variables is avoided when estimating the production frontier. Smoothing over time accounts for the panel structure of the data as well as the correlation over time of observed units.

The attainable set of pure inputs and output $(\varepsilon_x, \varepsilon_y)$ is defined by

$$\Psi_\varepsilon = \left\{ (e_x, e_y) \in \mathbb{R}^{p+1} \mid H_{\varepsilon_x, \varepsilon_y}(e_x, e_y) \geq 0 \right\} \tag{9.13}$$

where $H_{\varepsilon_x, \varepsilon_y}(e_x, e_y) = \Pr(\varepsilon_x \leq e_x, \varepsilon_y \geq e_y)$. The nonparametric FDH estimator is obtained by using the empirical analog $\widehat{H}_{\varepsilon_x, \varepsilon_y}(e_x, e_y)$ of $H_{\varepsilon_x, \varepsilon_y}(e_x, e_y)$ and the the observed residuals defined in (9.11) and (9.12). As shown in Florens et al. (2014), replacing the unobserved $(\varepsilon_x, \varepsilon_y)$ by their empirical counterparts $(\widehat{\varepsilon}_x, \widehat{\varepsilon}_y)$ does not change the usual statistical properties of frontier estimators. Consequently, both consistency for the full-frontier FDH estimator and $\sqrt{n}$-consistency and asymptotic normality for the robust order-m frontiers follow. Florens

[18] To some extent, the first step permits controlling for endogeneity due to reverse causation between production process of inputs and output and the external variables (Z, F_t).

et al. (2014) conjecture that if the functions μ_ℓ and σ_ℓ for $\ell = x, y$, are smooth enough, the conditional FDH estimator keeps its usual nonparametric rate of convergence i.e. here, $n^{1/(p+1)}$.

A "pure" measure of efficiency can be obtained by measuring the distance of a particular point $(\varepsilon_{x,it}, \varepsilon_{y,it})$ to the efficient frontier. Since the pure inputs and output are centered on zero, they may have negative values, requiring use of directional distance defined for a particular unit (e_x, e_y) by

$$\delta(e_x, e_y; d_x, d_y) = \sup\{\gamma | H_{\varepsilon_x,\varepsilon_y}(e_x - \gamma d_x, e_y + \gamma d_y) > 0\}, \tag{9.14}$$

where $d_x \in \mathbb{R}_+^p$ and $d_y \in \mathbb{R}_+$ are the chosen direction. In Mastromarco and Simar (2018) an output orientation is chosen by specifying $d_x = 0$ and $d_y = 1$. When only some elements of d_x are zero, see Daraio and Simar (2014) for details on practical computation.

In the output orientation and in the case of univariate output, the optimal production frontier can be described at any value of the pure input $e_x \in \mathbb{R}^p$ by the function

$$\varphi(e_x) = \sup\{e_y | H_{\varepsilon_x,\varepsilon_y}(e_x, e_y) > 0\}, \tag{9.15}$$

so that the distance to the frontier of a point (e_x, e_y), in the output direction is given directly by $\delta(e_x, e_y; 0, 1) = \varphi(e_x) - e_y$. Then, for each unit in the sample $\mathcal{X}_{n,s}$, the "pure" efficiency estimator is obtained through

$$\widehat{\delta}(\widehat{\varepsilon}_{x,it}, \widehat{\varepsilon}_{y,it}; 0, 1) = \widehat{\varphi}(\widehat{\varepsilon}_{x,it}) - \widehat{\varepsilon}_{y,it}, \tag{9.16}$$

where $\widehat{\varphi}(\cdot)$ is the FDH estimator of the pure efficient frontier in the output direction. The latter is obtained as

$$\begin{aligned} \widehat{\varphi}(e_x) &= \sup\{e_y | \widehat{H}_{\varepsilon_x,\varepsilon_y}(e_x, e_y) > 0\} \\ &= \max_{\{(i,t)|\widehat{\varepsilon}_{x,it} \leq e_x\}} \widehat{\varepsilon}_{y,it}. \end{aligned} \tag{9.17}$$

Similar expressions can be derived for the order-m efficiency estimator. The order-m frontier at an input value e_x, is the expected value of the maximum of the outputs of m units drawn at random in the population of units such that $\varepsilon_{x,it} \leq e_x$. The nonparametric estimator is obtained by the empirical analog

$$\widehat{\varphi}_m(e_x) = \widehat{E}\left[\max\left(\varepsilon_{y,1t}, \ldots, \varepsilon_{y,mt}\right)\right], \tag{9.18}$$

where the $\varepsilon_{y,it}$ are drawn from the empirical conditional survival function $\widehat{S}_{\varepsilon_y|\varepsilon_x}(e_y|\widehat{\varepsilon}_{x,it} \leq e_x)$. This can be computed by Monte-Carlo approximation or by solving a univariate integral using numerical methods (for practical details see Simar and Vanhems, 2012).

Note that it is possible to recover the conditional frontier in the original units, both for the full frontier and for the order-m one. As shown in Florens et al. (2014), it is directly obtained at any values of (x, z, f_t) as

$$\begin{aligned}\tau(x, z, f_t) &= \sup\{y|S_{Y|X,Z,F_t}(y|x, z, f_t) > 0\} \\ &= \mu_y(z, f_t) + \varphi(e_x)\sigma_y(z, f_t). \end{aligned} \tag{9.19}$$

which can be estimated by

$$\widehat{\tau}(X_{it}, Z_{it}, f_t) = \widehat{\mu}_y(Z_{it}, f_t) + \widehat{\varphi}(\widehat{\varepsilon}_{x,it})\widehat{\sigma}_y(Z_{it}, f_t). \tag{9.20}$$

As shown in Mastromarco and Simar (2018) this is equivalent to

$$\widehat{\tau}(X_{it}, Z_{it}, f_t) = Y_{it} + \widehat{\delta}(\widehat{\varepsilon}_{x,it}, \widehat{\varepsilon}_{y,it}; 0, 1)\widehat{\sigma}_y(Z_{it}, f_t), \tag{9.21}$$

which has the nice interpretation that gap, in original units, between an observation (X_{it}, Y_{it}) facing the conditions (Z_{it}, f_t) and the efficient frontier is given by the corresponding "pure" efficiency measure rescaled by the local standard deviation $\widehat{\sigma}_y(Z_{it}, f_t)$. In the same spirit, the conditional output oriented Farrell efficiency estimate in original units is given by

$$\widehat{\lambda}(X_{it}, Y_{it}|Z_{it}, f_t) = \frac{\widehat{\tau}(X_{it}, Z_{it}, f_t)}{Y_{it}}. \tag{9.22}$$

Note that there were no need to estimate the conditional survival function $S_{Y|X,Z,F_t}(y|x, z, f_t)$ to obtain this result. Similar results are described in Mastromarco and Simar (2018) or order-m robust frontiers.

Mastromarco and Simar (2018) have applied the approach in the analysis of the productivity performance of 26 OECD countries and 18 developing countries considering the spillover effects of global shocks and business cycles due to increasing globalization and interconnection among countries. So far all studies analyzing effect of common external factors on productivity of countries have been on the stream of parametric modeling which suffers of misspecification problem due to unknown data generation process in applied studies. The frontier model used by Mastromarco and Simar (2018) enables investigation of whether

the effect of environmental/global variables on productivity occurs via technology change or efficiency. Mastromarco and Simar (2018) quantify the impact of environmental/global factors on efficiency levels and make inferences about the contributions of these variables in affecting efficiency. Specifically, Mastromarco and Simar (2018) assess the impact of FDI on the production process for small, medium and large countries. They intend to redress an important policy issue of whether the protection-oriented policy will hamper the production efficiency through limiting FDI by explicitly analyzing the relationship between efficiency and openness factor FDI dependent on size of country.

Mastromarco and Simar (2018) show that, especially for medium and big countries, FDI appears to play an important role in accelerating the technological change (shifts in the frontier) but with a decreasing effect at large values of FDI. This result confirms the theoretical hypothesis that FDI leads to increase in productivity by spurring competition (Glass and Saggi, 1998). Moreover their findings reveal that knowledge embodied in FDI is transferred for technology externalities (shift of the frontier) (Cohen and Levinthal, 1989), supporting the evidence highlighting that lowering trade barriers have exerted a significantly positive effects on productivity (e.g., Borensztein et al., 1998 and Cameron et al., 2005).

In a further development in the panel data context, Mastromarco and Simar (2017) address the problem of endogeneity due to latent heterogeneity. They analyze the influence of human capital (measured as average years of education in the population) on the production process of a country by extending the instrumental nonparametric approach proposed by Simar et al. (2016). The latent factor which is identified is the part of human capital independent of the life expectancy in the countries. it appears that this latent factor can be empirically interpreted as innovation, quality of the institutions and the difference in property rights systems among countries (see Mastromarco and Simar, 2017 for details).

10 Summary and Conclusions

Benchmarking production performance is a fascinating, but difficult field because of the nonstandard and challenging statistical and econometric problems that are involved. Nonparametric methods for efficiency estimation bring together a wide variety of mathematical tools from mathematical statistics, econometrics, computer science, and operations research. As this survey indicates, a lot of progress has been made in recent years in establishing

statistical theory for efficiency estimators. The development of new CLTs for both unconditional and conditional efficiency estimators opens the door to testing hypotheses about the structure of production models and their evolution over time. The extension of existing results to panel data similarly opens new doors to handle dynamic situations. These and other issues continue to be addressed by the authors of this survey as well as others.

Although we have focused on the nonparametric, deterministic framework, other possibilities falling somewhere between this framework and the framework of fully parametric, stochastic frontiers. For example, Fan et al. (1996) propose semi-parametric estimation of a stochastic frontier. Kumbhakar et al. (2004) propose a local likelihood approach that requires distributional assumptions for the noise and efficiency processes, but does not require functional-form assumptions for the frontier. Parameters of the noise and efficiency distributions are estimated locally, and hence the method is almost fully nonparametric. In a similar vein, Simar et al. (2014) use moment-based methods to avoid specification of the efficiency process, and also avoid specifying the distribution of (the symmetric) noise process. Their method has the additional advantage that from a computational viewpoint, it is much easier to implement than the method of Kumbhakar et al. (2004). Simar and Zelenyuk (2011) discuss stochastic version of the FDH and DEA estimators. Kneip et al. (2015a) allow for measurement error, while Florens et al. (2018) allow for symmetric noise with unknown variance in a nonparametric framework. In yet another direction, Kuosmanen and Kortelainen (2012) introduce shape constraints in a semi-parametric framework for frontier estimation, but so far all the stochastic parts of the model are fully parametric. All of these create new issues and new directions for future research.

References

Adler, N. and B. Golany (2001), Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to western europe, *European Jouranl of Operational Research* 132, 260–273.

— (2007), PCA-DEA: Reducing the curse of dimensionality, in J. Zhu and W. D. Cook, eds., *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis*, New York: Springer Science+Business Media, LLC, pp. 139–154.

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.

Bahadur, R. R. and L. J. Savage (1956), The nonexistence of certain statistical procedures in nonparametric problems, *The Annals of Mathematical Statistics* 27, 1115–1122.

Bai, J. (2009), Panel data models with interactive fixed effects, *Econometrica* 77, 1229–1279.

Banker, R., A. Charnes, W. W. Cooper, J. Swarts, and D. A. Thomas (1989), An introduction to data envelopment analysis with some of its models and uses, *Research in Governmental and Nonprofit Accounting* 5, 125–165.

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.

Bellman, R. E. (1957), *Dynamic Programming*, Princeton, NJ: Princeton University Press.

Bickel, P. J. and D. A. Freedman (1981), Some asymptotic theory for the bootstrap, *Annals of Statistics* 9, 1196–1217.

Bickel, P. J. and A. Sakov (2008), On the choice of m in the m out of n bootstrap and confidence bounds for extrema, *Statistica Sinica* 18, 967–985.

Bogetoft, P. (1996), DEA on relaxed convexity assumptions, *Managment Science* 42, 457–465.

Bogetoft, P., J. M. Tama, and J. Tind (2000), Convex input and output projections of nonconvex production possibility sets, *Management Science* 46, 858–869.

Borensztein, E., J. D. Gregorio, and L.-W. Lee (1998), How does foreign direct investment affect economic growth?, *Journal of International Economics* 45, 115–135.

Boussofiane, A., R. G. Dyson, and E. Thanassoulis (1991), Applied data envelopment analysis, *European Journal of Operational Research* 52, 1–15.

Bowlin, W. F. (1987), Evaluating the efficiency of US Air Force real-property maintenance activities, *Journal of the Operational Research Society* 38, 127–135.

— (1998), Measuring performance: An introduction to data envelopment analysis (DEA), *Journal of Applied Behavior Analysis* 15, 3–27.

Boyd, S. and L. Vandenberghe (2004), *Convex Optimization*, New York: Cambridge University Press.

Briec, W., K. Kerstens, and P. Van den Eeckaut (2004), Non-convex technologies and cost functions: definitions, duality, and nonparametric tests of convextiy, *Journal of Economics* 81, 155–192.

Bădin, L., C. Daraio, and L. Simar (2010), Optimal bandwidth selection for conditional efficiency measures: A data-driven approach, *European Journal of Operational Research* 201, 633–664.

— (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.

Cameron, G., J. Proudman, and S. Redding (2005), Technological convergence, R & D, trade and productivity growth, *European Economic Review* 49, 775–807.

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

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

Charnes, A., W. W. Cooper, B. Golany, L. Seiford, and J. Stutz (1985), Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical productions functions, *Journal of Econometrics* 30, 91–107.

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

— (1981), Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through, *Management Science* 27, 668–697.

Cohen, W. and D. Levinthal (1989), Innovation and learning: two faces of R & D, *Economics Journal* 107, 139–149.

Cooper, W. W., S. Li, L. M. Seiford, and J. Zhu (2004), Sensitivity analysis in DEA, in W. W. Cooper, L. M. Seiford, and J. Zhu, eds., *Handbook on Data Envelopment Analysis*, chapter 3, New York: Kluwer Academic Publishers, Inc., pp. 75–98.

Cooper, W. W., L. M. Seiford, and K. Tone (2000), *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, Boston, MA: Kluwer Academic Publishers.

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

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. and L. Simar (2005), Introducing environmental variables in nonparametric frontier models: A probabilistic approach, *Journal of Productivity Analysis* 24, 93–121.

— (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.

Daraio, C., L. Simar, and P. W. Wilson (2018), Nonparametric estimation of efficiency in the presence of environmental variables, *Econometrics Journal* In press, doi: 10.1111/ectj.12103.

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.

Durstenfeld, R. (1964), Algorithm 235: Random permutation, *Communications of the ACM* 7, 420.

Dyson, R. G., R. Allen, A. S. Camanho, V. V. Podinovski, C. S. Sarrico, and E. A. Shale (2001), Pitfalls and protocalls in DEA, *European Journal of Operational Research* 132, 245–259.

Fan, Y., Q. Li, and A. Weersink (1996), Semiparametric estimation of stochastic production frontier models, *Journal of Business and Economic Statistics* 14, 460–468.

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

Färe, R. and S. Grosskopf (2004), *Efficiency and Productivity: New Directions*, Boston, MA: Kluwer Academic Publishers.

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

Färe, R., S. Grosskopf, and D. Margaritis (2008), Productivity and efficiency: Malmquist and more, in H. Fried, C. A. K. Lovell, and S. Schmidt, eds., *The Measurement of Productive Efficiency*, chapter 5, Oxford: Oxford University Press, 2nd edition, pp. 522–621.

Färe, R. and C. A. K. Lovell (1988), Aggregation and efficiency.

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

Fisher, R. A. and F. Yates (1948), *Statistical Tables for Biological, Agricultural and Medical Research*, London: Oliver and Boyd, 3rd edition.

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

— (2018), Stochastic frontier estimation with symmetric error. Discussion paper #2018/xx, Institut de Statistique Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Friedman, L. and Z. Sinuany-Stern (1998), Combining ranking scales and selecting variables in the DEA context: The case of industrial branches, *Computers and Operations Research* 25, 781–791.

Glass, A. J. and K. Saggi (1998), International technological transfer and technology gap, *Journal of Development Economics* 55, 369–398.

Golany, B. and Y. Roll (1989), An application procedure for DEA, *Omega* 17, 237–250.

Gray, H. L. and R. R. Schucany (1972), *The Generalized Jackknife Statistic*, New York: Marcel Decker, Inc.

Homburg, C. (2001), Using data envelopement analysis to benchmark activities, *International Journal of Production Economics* 73, 51–58.

Hsu, P. L. and H. Robbins (1947), Complete convergence and the law of large numbers, *Proceedings of the National Academy of Sciences of the United States of America* 33, 25–31.

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.

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.

Kumbhakar, S. C., B. U. Park, L. Simar, and E. G. Tsionas (2004), Nonparametric stochastic frontiers: A local likelihood approach. Discussion paper #0417, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Kuosmanen, T. and M. Kortelainen (2012), Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints, *Journal of Productivity Analysis* 38, 11–28.

Mastromarco, C. and L. Simar (2015), Effect of FDI and time on catching up: New insights from a conditional nonparametric frontier analysis, *Journal of Applied Econometrics* 30, 826–847.

— (2017), Cross-section dependence and latent heterogeneity to evaluate the impact of human capital on country performance. Discussion paper #2017/30, Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

— (2018), Globalization and productivity: A robust nonparametric world frontier analysis, *Economic Modelling* 69, 134–149.

Mouchart, M. and L. Simar (2002), Efficiency analysis of air controllers: First insights. Consulting report #0202, Institut de Statistique, Université Catholique de Louvain, Belgium.

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.

Pesaran, M. H. (2006), Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica* 74, 967–1012.

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

Raab, R. and R. Lichty (2002), Identifying sub-areas that comprise a greater metropolitan area: the criterion of country relative efficiency, *Journal of Regional Science* 42, 579–594.

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

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

Simar, L., I. Van Keilegom, and V. Zelenyuk (2014), 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 I. Van Keilegom (2016), Unobserved heterogeneity and endogeneity in nonparametric frontier estimation, *Journal of Econometrics* 190, 360–373.

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.

— (1999a), Some problems with the Ferrier/Hirschberg bootstrap idea, *Journal of Productivity Analysis* 11, 67–80.

— (1999b), Of course we can bootstrap DEA scores! But does it mean anything? Logic trumps wishful thinking, *Journal of Productivity Analysis* 11, 93–97.

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

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

— (2008), Statistical inference in nonparametric frontier models: Recent developments and perspectives, in H. O. Fried, C. A. K. Lovell, and S. S. Schmidt, eds., *The Measurement of Productive Efficiency*, chapter 4, Oxford: Oxford University Press, 2nd edition, pp. 421–521.

— (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.

— (2013), Estimation and inference in nonparametric frontier models: Recent developments and perspectives, *Foundations and Trends in Econometrics* 5, 183–337.

Simar, L. and V. Zelenyuk (2011), Stochastic FDH/DEA estimators for frontier analysis, *Journal of Productivity Analysis* 36, 1–20.

Spanos, A. (1999), *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*, Cambridge: Cambridge University Press.

Varian, H. R. (1978), *Microeconomic Analysis*, New York: W. W. Norton & Co.

Vassiloglou, M. and D. Giokas (1990), A study of the relative efficiency of bank branches: an application of data envelopment analysis, *Journal of the Operational Research Society* 41, 591–597.

Wheelock, D. C. and P. W. Wilson (2008), Non-parametric, unconditional quantile estimation for efficiency analysis with an application to Federal Reserve check processing operations, *Journal of Econometrics* 145, 209–225.

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.

Zervopoulos, P. D., F. Vargas, and G. Cheng (2012), Reconceptualizing the DEA bootstrap for improved estimations in the presence of small samples, in R. Banker, A. Emrouznejad, A. L. M. Lopes, and M. R. de Almeida, eds., *Data Envelopment Analysis: Theory and Applications*, chapter 27, Brazil: Proceedings of the 10th International Conference on DEA, pp. 226–232.