The Econometrics Journal

Econometrics Journal (2018), volume **21**, pp. 170–191.
doi: 10.1111/ectj.12103

Central limit theorems for conditional efficiency measures and tests of the 'separability' condition in non-parametric, two-stage models of production

CINZIA DARAIO[†], LÉOPOLD SIMAR[‡] AND PAUL W. WILSON[§,‖]

[†]*Department of Computer, Control and Management Engineering Antonio Ruberti (DIAG), University of Rome 'La Sapienza', Piazzale Aldo Moro 5, 00185, Rome, Italy.*
E-mail: daraio@dis.uniroma1.it

[‡]*Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium.*
E-mail: leopold.simar@uclouvain.be

[§]*Department of Economics, Clemson University, Clemson, SC 29634, USA.*
E-mail: pww@clemson.edu

[‖]*School of Computing, Division of Computer Science, 100 McAdams Hall, Clemson University, Clemson, SC 29634, USA.*

First version received: December 2016; final version accepted: September 2017

Summary In this paper, we demonstrate that standard central limit theorem (CLT) results do not hold for means of non-parametric, conditional efficiency estimators, and we provide new CLTs that permit applied researchers to make valid inference about mean conditional efficiency or to compare mean efficiency across groups of producers. The new CLTs are used to develop a test of the restrictive 'separability' condition that is necessary for second-stage regressions of efficiency estimates on environmental variables. We show that if this condition is violated, not only are second-stage regressions difficult to interpret and perhaps meaningless, but also first-stage, unconditional efficiency estimates are misleading. As such, the test developed here is of fundamental importance to applied researchers using non-parametric methods for efficiency estimation. The test is shown to be consistent and its local power is examined. Our simulation results indicate that our tests perform well both in terms of size and power. We provide a real-world empirical example by re-examining the paper by Aly et al. (1990, *Review of Economics and Statistics 72*, 211–18) and rejecting the separability assumption implicitly assumed by Aly et al., calling into question results that appear in hundreds of papers that have been published in recent years.

Keywords: *Conditional efficiency, Data envelopment analysis (DEA), Free-disposal hull (FDH), Separability, Technical efficiency, Two-stage estimation.*

1. INTRODUCTION

Non-parametric efficiency estimators are widely used to benchmark the performance of firms and other decision-making units. Unconditional versions of these estimators measure distance from a particular point in input–output space to an estimate of the boundary of the attainable set, i.e., the set of feasible combinations of inputs and outputs. Farrell (1957) is the first empirical example of such estimators, and relies on the convex hull of a set of observed input–output combinations to estimate the attainable set. This method has been popularized by Charnes et al. (1978) and is known in the literature as data envelopment analysis (DEA).[1] Deprins et al. (1984) relaxed the convexity assumption in the DEA estimator by using the free-disposal hull (FDH) of a set of observed input–output combinations to estimate the attainable set. More recently, Daraio and Simar (2005) have developed conditional measures of efficiency, allowing non-parametric estimation of technical efficiency conditional on some explanatory, contextual, 'environmental' variables that are neither inputs nor outputs in the production process. Recent surveys of both the unconditional and conditional estimators are provided by Simar and Wilson (2013, 2015).

Kneip et al. (2015) demonstrate that conventional central limit theorem (CLT) results do not hold for sample means of unconditional DEA and FDH estimates, and they provide new CLTs that enable inference about mean efficiency using asymptotic normal approximations. Kneip et al. (2016) extend these results to provide tests of (a) differences in mean efficiency across groups of producers, (b) convexity versus non-convexity of production sets, and (c) constant versus variable returns to scale, but only in the absence of environmental variables. In this paper, we show that conventional CLTs also do not hold for sample means of conditional DEA and FDH estimators, and we provide new CLTs that enable applied researchers to make inferences about mean conditional efficiency. On the surface, this extension parallels the development in Kneip et al. (2015), but some additional complication arises due to the presence of a bandwidth parameter in the conditional estimators. Because of the presence of bandwidths, new theoretical results on moments, CLTs, etc., are needed and these are provided for conditional efficiency estimators. In addition, we develop a statistical test of the restrictive 'separability' condition described by Simar and Wilson (2007) that must be maintained when DEA or FDH efficiency estimates are regressed on explanatory variables in a second-stage estimation exercise. Given the restrictiveness of the separability condition as discussed below, the continued proliferation of studies regressing DEA or FDH efficiency estimates on variables in a second stage, and the lack, to date, of any test of the separability condition, the development here of such a test is an important contribution. The test is shown to be consistent and its local power is examined.

The presence of environmental variables raises important questions for practitioners, such as the question of precisely how the environmental variables might affect the production process. Conceivably, the environmental variables might affect only the distribution of efficiency among firms. However, environmental variables might affect the production possibilities of firms, or environmental variables might affect both the distribution of efficiency as well as production possibilities. The separability condition described by Simar and Wilson (2007) amounts to an assumption that environmental variables only affect the distribution of efficiency and do not affect production possibilities. If the separability condition does not hold, unconditional DEA and FDH estimators have no useful interpretation; that is, not only are second-stage regressions difficult to interpret and perhaps without meaning when the separability condition is violated, but also the (unconditional) first-stage efficiency estimates do not estimate any meaningful model feature.

In the next section, we establish notation and develop the statistical framework. The requisite estimators are briefly discussed in Section 3. Our main theoretical results appear in

[1] Banker et al. (1984) modified the Farrell (1957) estimator by using the conical hull of a set of observed input–output combinations to estimate the attainable set, thereby imposing an assumption of constant returns to scale.

Sections 4 and 5. In Section 4, we first derive results on asymptotic moments of conditional efficiency estimators. We then use these results to show that standard CLTs (e.g., Lindeberg–Feller) do not hold for means of conditional efficiency estimates, and to develop new CLTs for means of conditional efficiency estimates. We also discuss in Section 4 the bias estimates and subsampling methods needed for the new CLT results. Then, in Section 5, we use the new CLTs to develop statistical tests of separability versus non-separability. In Section 6, we revisit the empirical work of Aly et al. (1990) who estimate efficiency for a subsample of US commercial banks, and then we regress the efficiency estimates on some environmental variables. In particular, our tests provide strong evidence that the separability condition implicitly assumed by Aly et al. (1990) does not hold, calling into question the results of their second-stage regression. Conclusions are given in the final section.

Separately, the online Appendices contain additional details. Online Appendix A gives technical assumptions used to derive results in Section 5, and proofs of lemmata and theorems are given in online Appendix B. In online Appendix C, we discuss how one can handle discrete environmental variables. The supplementary material mentioned in Section 5.6 appears in online Appendix D. The results of Monte Carlo experiments providing evidence on the size and power of the proposed tests, mentioned briefly in Section 5.7, appear in online Appendix E.

2. THE PRODUCTION PROCESS IN THE PRESENCE OF ENVIRONMENTAL FACTORS

In this section, we formalize a statistical model of the production process along the lines of the probability framework of Cazals et al. (2002). The production process generates random variables (X, Y, Z) in an appropriate probability space, where $X \in \mathbb{R}_+^p$ is the vector of input quantities, $Y \in \mathbb{R}_+^q$ is the vector of output quantities and $Z \in \mathbb{R}^r$ is a vector of variables describing environmental factors. The elements in Z are neither inputs nor outputs and are typically not under the control of the manager, but they may influence the production process in different ways as explained below. 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

$$f_{XYZ}(x, y, z) = f_{XY|Z}(x, y \mid z) f_Z(z). \tag{2.1}$$

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

$$\Psi^z = \{(X, Y) \mid X \text{ can produce } Y \text{ when } Z = z\}. \tag{2.2}$$

The environmental variables in Z can affect the production process in one of the following ways: (a) only through Ψ^z, the support of (X, Y); (b) only through the density $f_{XY|Z}(x, y \mid z)$, thereby affecting the probability for a firm to be near its optimal boundary; (c) through both Ψ^z and $f_{XY|Z}(x, y \mid z)$. Let

$$\Psi = \bigcup_{z \in \mathcal{Z}} \Psi^z. \tag{2.3}$$

By construction, $\Psi^z \subseteq \Psi \ \forall\, z \in \mathcal{Z}$, and clearly $\Psi \subset \mathbb{R}_+^{p+q}$. However, whether Ψ is useful for benchmarking the performance of a firm producing output levels y from input levels x while facing levels z of the environmental variables depends on whether the separability condition described by Simar and Wilson (2007) is satisfied. This condition requires that Z affect production only through the conditional density $f_{XY|Z}(x, y \mid z)$ without affecting its support Ψ^z, and is stated explicitly in Assumption 2.1.

ASSUMPTION 2.1. (SEPARABILITY CONDITION) $\Psi^z = \Psi$ *for all* $z \in \mathcal{Z}$.

Clearly, when Assumption 2.1 holds, the joint support of (X, Y, Z) can be factorized as

$$\mathcal{P} = \Psi \times \mathcal{Z}, \tag{2.4}$$

and Ψ can be interpreted as the unconditional attainable set

$$\Psi = \{(X, Y) \mid X \text{ can produce } Y\}. \tag{2.5}$$

However, Ψ has the interpretation in (2.5) if and only if (iff) Assumption 2.1 holds. The separability condition is very strong and restrictive. Under Assumption 2.1, the environmental factors influence neither the shape nor the level of the boundary of the attainable set, and the potential effect of Z on the production process is only through the distribution of the inefficiencies. If the separability condition holds, it is meaningful to measure the efficiency of a particular production plan (x, y) by its distance to the boundary of Ψ. For example, under separability, the output-oriented Farrell efficiency score is given by

$$\begin{aligned} \lambda(x, y) &= \sup\{\lambda > 0 \mid (x, \lambda y) \in \Psi\} \\ &= \sup\{\lambda > 0 \mid H_{XY}(x, \lambda y) > 0\}, \end{aligned} \tag{2.6}$$

where $H_{XY}(x, y) = \Pr(X \leq x,\ Y \geq y)$ is the probability of finding a firm dominating the production unit operating at the level (x, y).[2]

In this case, it is meaningful to analyse the behaviour of $\lambda(x, y)$ as a function of Z by using an appropriate regression model; see Simar and Wilson (2007, 2011) for details.[3]

Alternatively, if the separability condition does not hold, then we have a more general situation where Z may influence the level and the shape of the boundary of the attainable sets (and may also influence the conditional density $f_{XY|Z}(x, y \mid z)$). The following assumption characterizes this situation explicitly.

ASSUMPTION 2.2. (NON-SEPARABILITY ASSUMPTION) $\Psi^z \neq \Psi$ *for some* $z \in \mathcal{Z}$.

Note that Assumptions 2.1 and 2.2 are mutually exclusive; one and only one holds in a given situation. Under Assumption 2.2, the efficiency measure in (2.6) is difficult to interpret; in fact, it is economically meaningless because it does not measure the distance to the appropriate boundary. If Assumption 2.2 holds, the set Ψ can still be defined as in (2.3), but for benchmarking production units, the boundary of Ψ has little interest in this case because it may be unattainable for some firms faced with unfavourable conditions represented described by z. In such cases, the conditional measure

$$\begin{aligned} \lambda(x, y \mid z) &= \sup\{\lambda > 0 \mid (x, \lambda y) \in \Psi^z\} \\ &= \sup\{\lambda > 0 \mid H_{XY|Z}(x, \lambda y \mid z) > 0\}, \end{aligned} \tag{2.7}$$

[2] Note that, as usual, inequalities involving vectors are defined on an element-by-element basis.

[3] We focus the presentation in this paper using output-oriented measures of efficiency such as the one in (2.6), but of course efficiency can be measured in other directions as desired. See the recent surveys by Simar and Wilson (2013, 2015), and references therein, for details. All of the results here are easily generalized to input, hyperbolic, and directional distance functions after straightforward (but perhaps tedious) changes in notation.

introduced by Cazals et al. (2002) and Daraio and Simar (2005), where $H_{XY|Z}(x, y \mid z) = \Pr(X \leq x,\ Y \geq y \mid Z = z)$ is the probability of finding a firm dominating the production unit operating at the level (x, y) and facing environmental conditions z and is the distribution function corresponding to the conditional density $f_{XY|Z}(x, y \mid z)$ introduced earlier, gives a measure of distance to the appropriate, relevant boundary (i.e., the boundary that is attainable by firms operating under conditions described by z).

The distinction between Assumptions 2.1 and 2.2, and their implications for how environmental variables in Z affect the production process, has often been neglected in the literature where researchers analyse the effect of Z on $\lambda(X, Y)$ by estimating some regression of $\lambda(X, Y)$ on Z. Typically, starting with a sample of observations $\{(X_i, Y_i, Z_i)\}_{i=1}^n$, DEA or FDH estimators $\widehat{\lambda}(X_i, Y_i)$ computed in a first stage are regressed on Z_i in a second-stage analysis. Even if Assumption 2.1 holds, additional problems described in Simar and Wilson (2007) remain to be solved in the second stage to obtain sensible inference. Theoretical results on how to make inference in a second-stage linear regression, when appropriate, are described in detail by Kneip et al. (2015). However, if Assumption 2.2 holds, the two-stage approach is almost certain to lead to incorrect results and inferences about the effect of Z on the production process. This explains why it is important, as noted by Simar and Wilson (2007), indeed essential, to test Assumption 2.1 against Assumption 2.2. If the test rejects separability in favour of Assumption 2.2, then only a second-stage regression of the conditional measure $\lambda(X, Y \mid Z)$ on Z can be meaningful, as described for example in Bădin et al. (2012).

3. NON-PARAMETRIC EFFICIENCY ESTIMATORS

The literature on non-parametric statistical inference for efficiency scores is by now well developed; see Simar and Wilson (2013, 2015) for recent surveys. Here, we summarize the definitions and properties needed to test Assumption 2.1 versus Assumption 2.2. Consider a sample of independent and identically distributed (i.i.d.) observations $\mathcal{S}_n = \{(X_i, Y_i, Z_i) \mid i = 1, \ldots, n\}$. Following Deprins et al. (1984), the FDH estimator $\widehat{\lambda}_{\mathrm{FDH}}(x, y \mid \mathcal{S}_n)$ of $\lambda(x, y)$ is obtained by replacing Ψ in (2.6) with the FDH of $\mathcal{S}_n$, or alternatively by replacing $H_{XY}(x, y)$ with its empirical analogue

$$\widehat{H}_{XY}(x, y) = n^{-1} \sum_{i=1}^{n} I(X_i \leq x, Y_i \geq y), \tag{3.1}$$

where $I(\cdot)$ is the indicator function. Replacing Ψ with the convex hull of the FDH of $\mathcal{S}_n$ in (2.6) gives the DEA efficiency estimator $\widehat{\lambda}_{\mathrm{DEA}}(x, y \mid \mathcal{S}_n)$ of $\lambda(x, y)$. Note that in both of these unconditional estimators, the data on Z_i are ignored; only the first $(p + q)$ components of the ordered $(p + q + r)$-tuples in $\mathcal{S}_n$ are used.

For the conditional efficiency scores we need to use a smoothed estimator of $H_{XY|Z}(x, y \mid z)$ to plug in (2.7), which requires a vector of bandwidths for Z.[4] Denoting this r-vector of

[4] See Simar et al. (2016) to justify this approach.

bandwidths by h, the conditional distribution function $H_{XY|Z}(x, y \mid z)$ is replaced by the estimator

$$\widehat{H}_{XY|Z}(x, y \mid z) = \frac{\sum_{i=1}^{n} I(X_i \leq x, Y_i \geq y) K_h(Z_i - z)}{\sum_{i=1}^{n} K_h(Z_i - z)}, \tag{3.2}$$

where $K_h(\cdot) = (h_1, \ldots, h_r)^{-1} K((Z_i - z)/h)$ 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 conditional FDH estimator

$$\widehat{\lambda}_{\mathrm{FDH}}(x, y \mid z, \mathcal{S}_n) = \max_{i \in \mathcal{I}_{\mathrm{FDH}}(z,h)} \left(\min_{j=1,\ldots,p} \left(\frac{Y_i^j}{y^j} \right) \right), \tag{3.3}$$

where $\mathcal{I}_{\mathrm{FDH}}(z, h) = \{i \mid z - h \leq Z_i \leq z + h \ \cap \ X_i \leq x\}$.

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

$$\widehat{\lambda}_{\mathrm{DEA}}(x, y \mid z, \mathcal{S}_n) = \max_{\lambda, \omega_1, \ldots, \omega_n} \Big\{ \lambda > 0 \mid \lambda y \leq \sum_{i \in \mathcal{I}_{\mathrm{DEA}}(z,h)} \omega_i Y_i, \ x \geq \sum_{i \in \mathcal{I}_{\mathrm{DEA}}(z,h)} \omega_i X_i, \text{ for some } \omega_i \geq 0 \text{ such that } \sum_{i \in \mathcal{I}_{\mathrm{DEA}}(z,h)} \omega_i = 1, \Big\} \tag{3.4}$$

where $\mathcal{I}_{\mathrm{DEA}} = \{i \mid z - h \leq Z_i \leq z + h\}$. Note that the conditional estimators in (3.3) and (3.4) are just localized versions of the unconditional FDH and DEA efficiency estimators, where the degree of localization is controlled by the bandwidth in h. Practical aspects for choosing bandwidths are discussed below in Section 5.5.

The properties of non-parametric efficiency estimators have been examined in a number of papers in recent years; again, see Simar and Wilson (2013, 2015) for summaries and references. Under appropriate assumptions, both the unconditional FDH and DEA estimators are consistent and converge to limiting distributions at rates n^{κ} where $\kappa = 1/(p + q)$ in the FDH case or $\kappa = 2/(p + q + 1)$ in the DEA case. Kneip et al. (2015) provide results on the moments of both unconditional FDH and DEA estimators.[5] 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 (2011) 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 (2011).

Jeong et al. (2010) show that the conditional versions of the FDH and DEA efficiency estimators share properties similar to their unconditional counterparts whenever the elements of Z are continuous.[6] 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,

[5] Here and in the exposition that follows, we omit the subscripts 'FDH' and 'DEA' from the efficiency estimator in order to describe results in a generic fashion, thereby conserving space.

[6] We discuss in online Appendix C how discrete 'environmental' variables can be handled. Otherwise, except in online Appendix C, we assume throughout that all elements of Z are continuous.

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 $n \to \infty$,

$$n_h^{\kappa}(\widehat{\lambda}(x, y \mid z, \mathcal{S}_n) - \lambda(x, y \mid z)) \xrightarrow{\mathcal{L}} Q_{xy|z}(\cdot) \tag{3.5}$$

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 regularity conditions discussed in the next section, but also on the property that there are enough points in a neighbourhood 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 $n^{-\alpha}$ with $\alpha < 1/r$).

4. NEW RESULTS ON CONDITIONAL EFFICIENCY ESTIMATORS

4.1. *Asymptotic moments of conditional efficiency estimators*

As noted by Kneip et al. (2015), availability of the asymptotic results for efficiency estimated at a fixed point (x, y) is useful, but not sufficient for analysing the behaviour of statistics that are functions of FDH or DEA estimators evaluated at random points (X_i, Y_i). In the discussion below, we denote the FDH and DEA efficiency estimators by $\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n)$ to stress the fact that the estimator is to be evaluated at a random point (X_i, Y_i).

Kneip et al. (2015) prove that for the unconditional FDH and DEA estimators, under some regularity conditions (see Kneip et al., 2015 for details) and as $n \to \infty$,

$$E[\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n) - \lambda(X_i, Y_i)] = Cn^{-\kappa} + R_{n,\kappa} \tag{4.1}$$

$$E[(\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n) - \lambda(X_i, Y_i))^2] = o(n^{-\kappa}), \tag{4.2}$$

and

$$|\mathrm{Cov}(\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n) - \lambda(X_i, Y_i), \widehat{\lambda}(X_j, Y_j \mid \mathcal{S}_n) - \lambda(X_j, Y_j))| = o(n^{-1}) \tag{4.3}$$

for all $i, j \in \{1, \ldots, n\}$, $i \neq j$ and where $R_{n,k} = o(n^{-\kappa})$. The values of the constant C, the rate κ, and the remainder term $R_{n,\kappa}$ depend on which estimator is used. For the DEA estimator, $\kappa = 2/(p + q + 1)$ and $R_{n,\kappa} = O(n^{-3\kappa/2}(\log n)^{\alpha_1})$; for the FDH estimator, $\kappa = 1/(p + q)$ and $R_{n,\kappa} = O(n^{-2\kappa}(\log n)^{\alpha_2})$. The values of $\alpha_j > 1$, $j = 1, 2$ are given in Kneip et al. (2015). For purposes of the results needed here, the $\log n$ factor contained in $R_{n,\kappa}$ does not play a role and can be ignored. The results outlined here are valid under a set of corresponding regularity assumptions; see Theorems 3.1 and 3.3 in Kneip et al. (2015).

Similar results are needed for the asymptotic moments of the conditional efficiency estimators. Following Jeong et al. (2010), note that for a given h, the conditional FDH and DEA estimators in (3.3) and (3.4) do not target $\lambda(x, y \mid z)$, but instead estimate

$$\lambda^h(x, y \mid z) = \sup\{\lambda > 0 \mid (x, \lambda y) \in \Psi^{z,h}\}, \tag{4.4}$$

with the conditional attainable set given by

$$\begin{aligned}\Psi^{z,h} &= \{(X, Y) \mid X \text{ can produce } Y, \text{ when } |Z - z| \le h\} \\ &= \{(x, y) \in \mathbb{R}_+^{p+q} \mid H^h_{XY|Z}(x, y \mid z) > 0\} \\ &= \{(x, y) \in \mathbb{R}_+^{p+q} \mid f^h_{XY|Z}(\cdot, \cdot \mid z) > 0\}. \end{aligned} \tag{4.5}$$

Here, $H^h_{XY|Z}(x, y \mid z) = \Pr(X \le x, Y \ge y \mid z - h \le Z \le z + h)$ 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-neighbourhood of z and $f^h_{XY|Z}(\cdot, \cdot \mid z)$ is the corresponding conditional density of (X, Y) given $|Z - z| \le h$ implicitly defined by

$$H^h_{XY|Z}(x, y \mid z) = \int_{-\infty}^{x} \int_{y}^{\infty} f^h_{XY|Z}(u, v \mid Z \in [z - h, z + h])\, dv\, du. \tag{4.6}$$

Alternatively, (4.4) can be written as

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

Moreover, it is clear that $\Psi^{z,h} = \bigcup_{|\tilde{z}-z| \le h} \Psi^{\tilde{z}}$.

Consequently, for all points (x, y) in the support of $f_{XY|Z}(x, y \mid z)$, the error of estimation can be decomposed as

$$\widehat{\lambda}(x, y \mid z) - \lambda(x, y \mid z) = \underbrace{\widehat{\lambda}(x, y \mid z) - \lambda^h(x, y \mid z)}_{=\Delta_1} + \underbrace{\lambda^h(x, y \mid z) - \lambda(x, y \mid z)}_{=\Delta_2}, \tag{4.8}$$

where the first difference (Δ_1) is due to the estimation error in the localized problem and the second difference (Δ_2) is the non-random discrepancy (≤ 0) introduced by the localization.

Some assumptions are needed to define a statistical model. The next three assumptions are conditional analogues of standard assumptions made by Shephard (1970), Färe (1988), Kneip et al. (2015) and others.

ASSUMPTION 4.1. *For all $z \in \mathcal{Z}$, Ψ^z and $\Psi^{z,h}$ are closed.*

ASSUMPTION 4.2. *For all $z \in \mathcal{Z}$, both inputs and outputs are strongly disposable; i.e., for any $z \in \mathcal{Z}$, $\widetilde{x} \ge x$ and $0 \le \widetilde{y} \le y$, 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 4.2 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.

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 4.3. *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 neighbourhood of the frontier of Ψ^z.*

ASSUMPTION 4.4. *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 4.4 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).

A number of additional assumptions are needed to complete the statistical model and to permit statistical analysis of the conditional estimators that have been introduced above as well as the test statistics introduced below. These assumptions are given in online Appendix A. Depending on the estimators that are used in a particular situation (i.e., either DEA or FDH), only a subset of the assumptions listed in online Appendix A are required.

Note that if Z is separable and has no effect on the frontier, then Assumption 4.4 is trivially satisfied for all h. As noted in Bădin et al. (2017), it is easy to show that if $h \propto n^{-\gamma}$ with $1/r > \gamma > 1/(r + \kappa^{-1})$, the difference in Assumption 4.4 will be $o(n_h^{-\kappa})$. We need $\gamma < r^{-1}$ to ensure there are enough observations in the h-neighbourhood of z; see Proposition 1 in Jeong et al. (2010). As we cannot find an explicit expression for the second component Δ_2 in (4.8), and as the Weibull distribution linked to the first component Δ_1 contains unknown parameters, the best that can be done is to determine the order of an optimal bandwidth by balancing the order of the two error terms, which leads to $h \propto n^{-1/(r+\kappa^{-1})}$, and then, as is usual in nonparametric smoothing techniques, to take a smaller bandwidth to eliminate the bias term due to the localization, as suggested in Assumption 2 of Jeong et al. (2010). As expected, the order of the optimal bandwidth depends on the dimensions of Z as well as of X and Y. In Section 5.5, we show how to select bandwidths h of appropriate order in applied work (see also the discussions in Bădin et al., 2017).

The following result provides moments for the conditional efficiency estimators.

THEOREM 4.1. *Suppose Assumptions 4.1, 4.2, 4.3, A.1, A.2 and A.4 hold. Then, under Assumption A.5 for the FDH case, or under Assumptions A.3 and A.6 for the DEA case, as* $n \to \infty$,

$$E\left[\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_n) - \lambda^h(X_i, Y_i \mid Z_i)\right] = C_c n_h^{-\kappa} + R_{c,n_h,\kappa}, \tag{4.9}$$

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

$$E\left[(\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_n) - \lambda^h(X_i, Y_i \mid Z_i))^2\right] = o(n_h^{-\kappa}), \tag{4.10}$$

and

$$|\mathrm{Cov}(\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_n) - \lambda^h(X_i, Y_i \mid Z_i),\ \widehat{\lambda}(X_j, Y_j \mid Z_j, \mathcal{S}_n) - \lambda^h(X_j, Y_j \mid Z_j))| = o\left(n_h^{-1}\right) \tag{4.11}$$

for all $i, j \in \{1, \ldots, n\}$, $i \neq j$. *In addition, for the conditional DEA estimator* $R_{c,n_h,\kappa} = O(n_h^{-3\kappa/2}(\log n_h)^{\alpha_1})$ *and for the conditional FDH estimator* $R_{c,n_h,\kappa} = O(n_h^{-2\kappa}(\log n_h)^{\alpha_2})$.

The role of the bandwidths required for the conditional estimators becomes apparent in Theorem 4.1. In particular, the bandwidths reduce the effective number of observations used to estimate the moments as the rates n^κ for the unconditional estimators are reduced to $n_h^\kappa = n^{\kappa/(\kappa r+1)}$ for the conditional estimators. As will be seen, the $\log(n_h)$ factors appearing in the expressions for $R_{c,n_h,\kappa}$ do not play a role in the results that are derived below.

4.2. Central limit theorems for conditional efficiency estimators

Consider the sample means

$$\widehat{\mu}_n = n^{-1} \sum_{i=1}^{n} \widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n) \tag{4.12}$$

and

$$\widehat{\mu}_{c,n} = n^{-1} \sum_{i=1}^{n} \widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_n) \tag{4.13}$$

of unconditional and conditional efficiency estimators. The efficiency estimators in (4.12) and (4.13) could be either FDH or DEA estimators; differences between the two are noted below when relevant. In this subsection, we use the properties of moments of the conditional efficiency estimators derived in Section 4.1 to develop CLTs for means of conditional efficiency estimators.

For the case of means of unconditional efficiency estimators, Theorem 4.1 of Kneip et al. (2015) establishes that

$$\sqrt{n}(\widehat{\mu}_n - \mu - Cn^{-\kappa} - R_{n,\kappa}) \xrightarrow{\mathcal{L}} N(0, \sigma^2) \tag{4.14}$$

as $n \to \infty$, where $\mu = E[\lambda(X, Y)]$ and $\sigma^2 = \mathrm{Var}(\lambda(X, Y))$. The theorem also establishes that $\widehat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} (\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_n) - \widehat{\mu}_n)^2$ is a consistent estimator of σ^2. Conventional CLTs (e.g., the Lindeberg–Feller CLT) do not account for the bias term $Cn^{-\kappa}$, and hence are invalid for means of unconditional efficiency estimators unless $\kappa > 1/2$. In the case of FDH estimators, $\kappa > 1/2$ iff $(p + q) \leq 1$; in the case of DEA estimators, $\kappa > 1/2$ iff $(p + q) \leq 2$. If $\kappa = 1/2$, then the bias is stable as $n \to \infty$, but if $\kappa < 1/2$, the bias explodes asymptotically. Kneip et al. (2015) solve this problem by incorporating a generalized jackknife estimate of the bias and considering, when needed, test statistics based on averages over a subsample of observations. We use a similar approach below, although with the conditional efficiency estimators, the problem is more complicated than the one in Kneip et al. (2015) due to the localization in the conditional efficiency estimators.

Define

$$\mu_c^h = E[\lambda^h(X, Y \mid Z)] = \int_{\mathcal{P}} \lambda^h(x, y \mid z) f_{XYZ}(x, y, z)\, dx\, dy\, dz \tag{4.15}$$

and

$$\sigma_c^{2,h} = \mathrm{Var}(\lambda^h(X, Y \mid Z)) = \int_{\mathcal{P}} \left(\lambda^h(x, y \mid z) - \mu_c^h\right)^2 f_{XYZ}(x, y, z)\, dx\, dy\, dz, \tag{4.16}$$

where $\mathcal{P}$ is defined just before (2.1). These are the localized analogues 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

$$\frac{\sqrt{n}(\overline{\mu}_{c,n} - \mu_c^h)}{\sqrt{\sigma_c^{2,h}}} \xrightarrow{\mathcal{L}} N(0, 1) \tag{4.17}$$

under mild assumptions.

An obvious solution might be to estimate μ_c^h by $\widehat{\mu}_{c,n}$, but this proves to be problematic. To see this, define $\zeta_n = \widehat{\mu}_{c,n} - \overline{\mu}_{c,n}$. It is clear that $E[\zeta_n] = C_c n_h^{-\kappa} + R_{c,n_h,\kappa}$ by (4.9), and

$\text{Var}(\zeta_n) = o(n_h^{-1})$ due to (4.10) and (4.11). It follows that $\zeta_n - E[\zeta_n] = o_p(n_h^{-1/2})$. Now define $\widetilde{\mu}_{c,n} = E[\widehat{\mu}_{c,n}]$. Then

$$\widetilde{\mu}_{c,n} = \mu_c^h + C_c n_h^{-\kappa} + R_{c,n_h,\kappa}, \tag{4.18}$$

and it follows that

$$\begin{aligned} \widehat{\mu}_{c,n} - \widetilde{\mu}_{c,n} &= \overline{\mu}_{c,n} - \mu_c^h + \zeta_n - E[\zeta_n], \\ &= \overline{\mu}_{c,n} - \mu_c^h + o_p(n_h^{-1/2}). \end{aligned} \tag{4.19}$$

Clearly, $\sqrt{n}(\widehat{\mu}_{c,n} - \widetilde{\mu}_{c,n})/\sqrt{\sigma_c^{2,h}}$ diverges as $n \to \infty$ because, although $\sqrt{n}(\overline{\mu}_{c,n} - \mu_c^h) = O_p(1)$, $n^{1/2} o_p(n_h^{-1/2})$ diverges as $n_h = n^{1-\gamma r}$ with $1/(r + \kappa^{-1}) < \gamma < 1/r$. Changing the scaling and considering $n^a(\widehat{\mu}_{c,n} - \widetilde{\mu}_{c,n})$ for some a such that $0 < a < (1 - \gamma r)/2 < 1/2$ does not work because the limiting distribution collapses to a point mass at zero in this case. Consequently, it seems there is no way to develop a CLT for means of conditional efficiency estimators analogous to the one in (4.14) for means of unconditional efficiency estimators.

The following result is useful for the results developed below.

LEMMA 4.1. *Under the assumptions of Theorem 4.1, for $\kappa = 1/(p+q)$ in the case of the FDH estimator and for $\kappa = 2/(p+q+1)$ in the case of the DEA estimator,*

$$E[\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_n)] = \mu_c^h + C_c n_h^{-\kappa} + R_{c,n_h,\kappa} \tag{4.20}$$

and

$$\text{Var}(\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_n)) = \sigma_c^{2,h} + o(n_h^{-\kappa/2}), \tag{4.21}$$

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

Next, consider a random subsample $\mathcal{S}^*_{n_h}$ from $\mathcal{S}_n$ 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

$$\overline{\mu}_{c,n_h} = \frac{1}{n_h} \sum_{(X_i,Y_i,Z_i)\in\mathcal{S}^*_{n_h}} \lambda^h(X_i, Y_i \mid Z_i), \tag{4.22}$$

$$\widehat{\mu}_{c,n_h} = \frac{1}{n_h} \sum_{\{(X_i,Y_i,Z_i)\in\mathcal{S}^*_{n_h}\}} \widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_n), \tag{4.23}$$

and let $\widetilde{\mu}_{c,n_h} = E[\widehat{\mu}_{c,n_h}]$. Note that the estimators on the right-hand side of (4.23) are computed relative to the full sample $\mathcal{S}_n$, but the summation is over elements of the subsample $\mathcal{S}^*_{n_h}$.

The next result provides our first CLT for means of conditional efficiency estimators.

THEOREM 4.2. *Under the assumptions of Theorem 4.1, the following conditions hold as $n \to \infty$ with $\kappa = 1/(p+q)$ for the FDH case and $\kappa = 2/(p+q+1)$ for the DEA case: (a) $\widetilde{\mu}_{c,n_h} = \mu_c^h + C_c n_h^{-\kappa} + R_{c,n_h,\kappa}$; (b) $\widehat{\mu}_{c,n_h} - \widetilde{\mu}_{c,n_h} = \overline{\mu}_{c,n_h} - \mu_c^h + o_p(n_h^{-1/2})$; (c) $\sqrt{n_h}(\widehat{\mu}_{c,n_h} - \mu_c^h - C_c n_h^{-\kappa} - R_{c,n_h,\kappa})/\sqrt{\sigma_c^{2,h}} \xrightarrow{\mathcal{L}} N(0,1)$; (d) for $\widehat{\sigma}_{c,n}^{2,h} = n^{-1}\sum_{i=1}^n (\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_n) - \widehat{\mu}_{c,n})^2$, $\widehat{\sigma}_{c,n}^{2,h}/\sigma_c^{2,h} \xrightarrow{p} 1$.*

There are no cases where standard CLTs with rate $\sqrt{n}$ 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 2.1 holds). Theorem 4.2 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 $\kappa = 1/2$, then the bias term $C_c n_h^{-\kappa}$ does not vanish, and if $\kappa < 1/2$, the bias term explodes as $n \to \infty$. These cases are addressed below.

4.3. Bias corrections and subsample averaging

For the unconditional case, all necessary details can be found in Theorems 4.3 and 4.4 of Kneip et al. (2015). Here, we derive corresponding results for conditional efficiency estimators. Assume that the observations in $\mathcal{S}_n$ are randomly ordered, and to simplify notation, assume that n is even. Let $\mathcal{S}_{n/2}^{(1)}$ denote the set of the first $n/2$ observations from $\mathcal{S}_n$, and let $\mathcal{S}_{n/2}^{(2)}$ denote the set of remaining $n/2$ observations from $\mathcal{S}_n$.[7] Next, for $j \in \{1,\ 2\}$ define

$$\widehat{\mu}_{c,n/2}^{j} = (n/2)^{-1} \sum_{(X_i, Y_i, Z_i) \in \mathcal{S}_{n/2}^{(j)}} \widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_{n/2}^{(j)}). \tag{4.24}$$

Let $\widetilde{\mu}_{c,n/2} = E[\widehat{\mu}_{c,n/2}^{1}] = E[\widehat{\mu}_{c,n/2}^{2}]$ and define

$$\overline{\mu}_{c,n/2}^{j} = \frac{2}{n} \sum_{(X_i, Y_i, Z_i) \in \mathcal{S}_{n/2}^{(j)}} \lambda^h(X_i, Y_i \mid Z_i). \tag{4.25}$$

By (4.19),

$$\widehat{\mu}_{c,n/2}^{j} - \widetilde{\mu}_{c,n/2} = \overline{\mu}_{c,n/2}^{j} - \mu_c^h + o_p(n_h^{-1/2}) \tag{4.26}$$

for $j \in \{1,\ 2\}$. Now define $\widehat{\mu}_{c,n/2}^{*} = (\widehat{\mu}_{c,n/2}^{1} + \widehat{\mu}_{c,n/2}^{2})/2$. Clearly,

$$\widehat{\mu}_{c,n/2}^{*} - \widetilde{\mu}_{c,n/2} = \overline{\mu}_{c,n} - \mu_c^h + o_p(n_h^{-1/2}). \tag{4.27}$$

Subtracting (4.19) from (4.27) and rearranging terms yields

$$\widehat{\mu}_{c,n/2}^{*} - \widehat{\mu}_{c,n} = \widetilde{\mu}_{c,n/2} - \widetilde{\mu}_{c,n} + o_p(n_h^{-1/2}). \tag{4.28}$$

From (4.18) we have $\widetilde{\mu}_{c,n/2} - \widetilde{\mu}_{c,n} = C_c(2^\kappa - 1)n_h^{-\kappa} + R_{c,n_h,\kappa}^{*}$ where $R_{c,n_h,\kappa}^{*} = R_{c,n_h/2,\kappa} - R_{c,n_h,\kappa}$, yielding an estimator

$$\widetilde{B}_{\kappa,n_h}^{c} = (2^\kappa - 1)^{-1}(\widehat{\mu}_{c,n/2}^{*} - \widehat{\mu}_{c,n}) = C_c n_h^{-\kappa} + R_{c,n_h,\kappa}^{*} + o_p(n_h^{-1/2}) \tag{4.29}$$

of the leading bias term $C_c n_h^{-\kappa}$ in Theorem 4.2(c). Note that the remainder $R_{c,n_h,\kappa}^{*}$ has the same order $o(n_h^{-\kappa})$ as $R_{c,n_h,\kappa}$ and hence can be neglected.

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 $\widetilde{B}_{\kappa,n_h}^{c}$ can be reduced by repeating the above steps $K \ll \binom{n}{n/2}$ times,

[7] If n is odd, then $\mathcal{S}_{n/2}^{(1)}$ can contain the first $\lfloor n/2 \rfloor$ observations and $\mathcal{S}_{n/2}^{(2)}$ can contain the remaining $n - \lfloor n/2 \rfloor$ observations from $\mathcal{S}_n$. The fact that $\mathcal{S}_{n/2}^{(2)}$ contains one more observation than $\mathcal{S}_{n/2}^{(1)}$ makes no difference asymptotically.

shuffling the observations before each split of $\mathcal{S}_n$, and then averaging the bias estimates. This yields a generalized jackknife estimate

$$\widehat{B}^c_{\kappa,n_h} = K^{-1} \sum_{k=1}^{K} \widetilde{B}^c_{\kappa,n_h,k}, \tag{4.30}$$

where $\widetilde{B}^c_{\kappa,n_h,k}$ represents the value computed from (4.29) using the kth sample split.

Combining results yields the following.

THEOREM 4.3. *Under the Assumptions of Theorem 4.1, with $\kappa = 1/(p+q) \geq 1/3$ in the FDH case or $\kappa = 2/(p+q+1) \geq 2/5$ in the DEA case,*

$$\frac{\sqrt{n_h}\left(\widehat{\mu}_{c,n_h} - \mu^h_c - \widehat{B}^c_{\kappa,n_h} - R^*_{c,n_h,\kappa}\right)}{\sqrt{\sigma^{2,h}_c}} \xrightarrow{\mathcal{L}} N(0,1) \tag{4.31}$$

as $n \to \infty$.

If $\kappa < 1/3$ in the FDH case, or $\kappa < 2/5$ in the DEA case, the remainder term does not vanish fast enough and $\sqrt{n_h} R^*_{c,n_h,\kappa} \to \infty$ as $n \to \infty$. In such cases, efficiency scores must be averaged over a subsample of smaller size as in Kneip et al. (2015).

Define $n_{h,\kappa} = \lfloor n_h^{2\kappa} \rfloor$ so that $\sqrt{n_{h,\kappa}} < n_h^{1/2}$ when $\kappa < 1/2$. Then define

$$\widehat{\mu}_{c,n_{h,\kappa}} = \frac{1}{n_{h,\kappa}} \sum_{(X_i,Y_i,Z_i)\in\mathcal{S}^{**}_{n_{h,\kappa}}} \widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_n) \tag{4.32}$$

where $\mathcal{S}^{**}_{n_{h,\kappa}}$ is a random subsample of size $n_{h,\kappa}$ from $\mathcal{S}_n$.

THEOREM 4.4. *Under the Assumptions of Theorem 4.1, with $\kappa = 1/(p+q)$ in the FDH case or $\kappa = 2/(p+q+1)$ in the DEA case,*

$$\frac{\sqrt{n_{h,\kappa}}(\widehat{\mu}_{c,n_{h,\kappa}} - \mu^h_c - \widehat{B}^c_{\kappa,n_h} - R^*_{c,n_h,\kappa})}{\sqrt{\sigma^{2,h}_c}} \xrightarrow{\mathcal{L}} N(0,1), \tag{4.33}$$

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

REMARK 4.1. Kneip et al. (2015) note that for selected values of $p+q$, two different CLTs are available for means of unconditional efficiency estimators. The same is true for the conditional cases. With the DEA estimator when $p+q=4$ (so that $\kappa = 2/5$), using Theorem 4.3 neglects a term $\sqrt{n_h} R^*_{c,n_h,\kappa} = O(n_h^{-1/10})$, whereas using Theorem 4.4, and an average over a subsample we neglect a term $\sqrt{n_{h,\kappa}} R^*_{c,n_h,\kappa} = O(n_h^{-1/5})$ and we might expect a better approximation. For the conditional FDH estimator when $p+q=3$ (and hence $\kappa = 1/3$), using Theorem 4.3 implies an error of order $O(n_h^{-1/6})$, and using an average over a subsample implies, by Theorem 4.4, an error of the smaller order $O(n_h^{-1/3})$.

5. TESTING SEPARABILITY

5.1. Basic ideas

The goal is to test the null hypothesis of separability (Assumption 2.1) against its complement (Assumption 2.2). 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 2.1, $\lambda(X, Y) = \lambda(X, Y \mid Z)$ 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. However, under Assumption 2.2, the conditional attainable sets Ψ^z are different and the two estimators converge to different objects. Moreover, under Assumption 2.2, $\lambda(X, Y) \geq \lambda(X, Y \mid Z)$ with strict inequality holding for some $(X, Y, Z) \in \mathcal{P}$.

The approach developed here is similar to those developed in Kneip et al. (2016) for testing constant versus variable returns to scale or for testing convexity versus non-convexity of the attainable set. Recall the sample means in (4.12) and (4.13), where the efficiency estimators on the right-hand side of (4.12) and (4.13) could be either FDH or DEA estimators. For purposes of the following discussion, suppose the same type of estimators (FDH or DEA) are used in both (4.12) and (4.13). By construction $(\widehat{\mu}_n - \widehat{\mu}_{c,n}) \geq 0$, and the null hypothesis of separability should be rejected if this difference is too big. However, several problems remain to be solved, requiring some preliminary steps to adapt the existing results to the set-up here. We demonstrate in online Appendix E that the procedure works well in practice with finite sample sizes.

5.2. Test statistics

Let $\mu = E[\lambda(X, Y)]$, $\mu_c = E[\lambda(X, Y \mid Z)]$, and $\xi = \mu - \mu_c$. In addition, let $\xi^h = \mu - \mu_c^h$ where μ_c^h is defined in (4.15). Note that because for all h, $\mu \geq \mu_c^h \geq \mu_c$ and hence $\xi \geq \xi^h \geq 0$. Moreover, under the null $\xi^h = \xi = 0$ for all h. As noted above, in order to test the hypothesis that Z is separable, i.e., to test H_0 : Assumption 2.1 holds versus H_1 : Assumption 2.2 holds, one might consider the difference between estimators μ and μ_c^h, which under the null estimate the same quantity. When the null is true, $\lambda(X, Y) \equiv \lambda^h(X, Y|Z)$ with probability one, for all values of h. Under the null, the two estimators $\widehat{\mu}_n$ and $\widehat{\mu}_{c,n_h}$ have (when appropriately rescaled, depending on the value of κ), an asymptotic normal distribution with mean $\mu = \mu_c^h$ and variance $\sigma^2 = \sigma_c^{2,h}$ for all h, and so both are consistent estimators of the common μ. As explained in the preceding section, we can also, in both cases, correct for the inherent bias of the estimators. Clearly, the test can be stated equivalently in terms of $H_0 : \xi = 0$ versus $H_1 : \xi > 0$ for the reasons discussed in Section 5.1, indicating that a one-sided test is appropriate.

However, the properties of $(\widehat{\mu}_n - \widehat{\mu}_{c,n_h})$ (and their bias-corrected versions) are complicated because of the covariance between the two estimators, and this covariance is hard to estimate. Even in the limiting case where h is big enough so that $n_h = n$, it is clear that under the null, the asymptotic distribution of $(\widehat{\mu}_n - \widehat{\mu}_{c,n_h})$ will be degenerate with mass one at zero.[8]

The solution used here is analogous to the method used in the test for convexity of Ψ described by Kneip et al. (2016). In particular, the sample $\mathcal{S}_n$ can be split into two

[8] As observed by Hall et al. (2004), if Z is irrelevant in the production process (independent of (X, Y)), the optimal value of the bandwidth is infinity. This limiting case is more restrictive that the hypothesis to be tested here, but may arise in practice.

independent, parts, $\mathcal{S}_{1,n_1}$ and $\mathcal{S}_{2,n_2}$, 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.[9]

After splitting the sample, compute the estimators

$$\widehat{\mu}_{n_1} = n_1^{-1} \sum_{(X_i,Y_i)\in\mathcal{S}_{1,n_1}} \widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_{1,n_1}) \tag{5.1}$$

and

$$\widehat{\mu}_{c,n_{2,h}} = n_{2,h}^{-1} \sum_{(X_i,Y_i,Z_i)\in\mathcal{S}^*_{2,n_{2,h}}} \widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_{2,n_2}), \tag{5.2}$$

where, as in Section 4.2, $\mathcal{S}^*_{2,n_{2,h}}$ in (5.2) 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 samples by

$$\widehat{\sigma}^2_{n_1} = n_1^{-1} \sum_{(X_i,Y_i)\in\mathcal{S}_{1,n_1}} (\widehat{\lambda}(X_i, Y_i \mid \mathcal{S}_{1,n_1}) - \widehat{\mu}_{n_1})^2 \tag{5.3}$$

and

$$\widehat{\sigma}^{2,h}_{c,n_2} = n_2^{-1} \sum_{(X_i,Y_i,Z_i)\in\mathcal{S}_{2,n_2}} (\widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}_{2,n_2}) - \widehat{\mu}_{c,n_2})^2 \tag{5.4}$$

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

The estimators of bias corresponding to (4.29) for a single split of each subsample for the unconditional and conditional cases are given by

$$\widetilde{B}_{\kappa,n_1} = (2^\kappa - 1)^{-1}(\widehat{\mu}^*_{n_1/2} - \widehat{\mu}_{n_1}) \tag{5.5}$$

and

$$\widetilde{B}^c_{\kappa,n_{2,h}} = (2^\kappa - 1)^{-1}(\widehat{\mu}^*_{c,n_2/2} - \widehat{\mu}_{c,n_2}). \tag{5.6}$$

For the unconditional case in (5.5), $\widehat{\mu}^*_{n_1/2} = (\widehat{\mu}^1_{n_1/2} + \widehat{\mu}^2_{n_1/2})/2$, and for $j \in \{1,\ 2\}$, $\widehat{\mu}^j_{n_1/2} = (n_1/2)^{-1} \sum_{(X_i,Y_i,Z_i)\in\mathcal{S}^{(j)}_{n_1/2}} \widehat{\lambda}(X_i, Y_i \mid \mathcal{S}^{(j)}_{n_1/2})$, where $\mathcal{S}^{(j)}_{n_1/2}$ is the jth part of a random split of the full (sub)sample $\mathcal{S}_{n_1}$. Details are given in Kneip et al. (2015). For the conditional case in (5.6), $\widehat{\mu}^*_{c,n_2/2} = (\widehat{\mu}^1_{c,n_2/2} + \widehat{\mu}^2_{c,n_2/2})/2$, and for $j \in \{1,\ 2\}$, $\widehat{\mu}^j_{c,n_2/2} = (n_2/2)^{-1} \sum_{(X_i,Y_i,Z_i)\in\mathcal{S}^{(j)}_{n_2/2}} \widehat{\lambda}(X_i, Y_i \mid Z_i, \mathcal{S}^{(j)}_{n_2/2})$, where $\mathcal{S}^{(j)}_{n_2/2}$ is the jth part of a random split of the full (sub)sample $\mathcal{S}_{n_2}$. The bias

[9] 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 the order of $n^{-\kappa/(r\kappa+1)}$, 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 and 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 $n_2 = 178$.

estimates in (5.5)–(5.6) can then be averaged over K random splits of the two subsamples $\mathcal{S}_{n_1}$ and $\mathcal{S}_{n_2}$ to obtain bias estimates $\widehat{B}_{\kappa,n_1}$ for the unconditional case and $\widehat{B}^c_{\kappa,n_{2,h}}$ for the conditional case.

For small values of $(p+q)$ such that $\kappa \geq 1/3$ in the FDH case or $\kappa \geq 2/5$ when DEA estimators are used, Theorem 4.3 and Theorem 4.3 of Kneip et al. (2015) can be used to construct an asymptotically normal test statistic for testing the null hypothesis of separability. In particular, because our bias-corrected sample means are independent due to splitting the original sample into independent parts, and because 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

$$T_{1,n} = \frac{(\widehat{\mu}_{n_1} - \widehat{\mu}_{c,n_{2,h}}) - (\widehat{B}_{\kappa,n_1} - \widehat{B}^c_{\kappa,n_{2,h}})}{\sqrt{(\widehat{\sigma}^2_{n_1}/n_1) + (\widehat{\sigma}^{2,h}_{c,n_2}/n_{2,h})}} \xrightarrow{\mathcal{L}} N(0,1) \tag{5.7}$$

under the null. Alternatively, for $\kappa < 1/2$, similar reasoning with Theorem 4.4 and Theorem 4.4 of Kneip et al. (2015) leads to

$$T_{2,n} = \frac{(\widehat{\mu}_{n_{1,\kappa}} - \widehat{\mu}_{c,n_{2,h,\kappa}}) - (\widehat{B}_{\kappa,n_1} - \widehat{B}^c_{\kappa,n_{2,h}})}{\sqrt{(\widehat{\sigma}^2_{n_1}/n_{1,\kappa}) + (\widehat{\sigma}^{2,h}_{c,n_2}/n_{2,h,\kappa})}} \xrightarrow{\mathcal{L}} N(0,1) \tag{5.8}$$

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. (2015) 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 $\mathcal{S}_n$, we can compute values $\widehat{T}_{1,n}$ or $\widehat{T}_{2,n}$ depending on the value of $(p+q)$. The null should be rejected whenever $1 - \Phi(\widehat{T}_{1,n})$ or $1 - \Phi(\widehat{T}_{2,n})$ is less than the desired test size, e.g., 0.1, 0.05, or 0.01, where $\Phi(\cdot)$ denotes the standard normal distribution function.

5.3. *Consistency of the test*

It is easy to show that tests based on $T_{1,n}$ or $T_{2,n}$ (as appropriate, depending on the value of κ) are consistent. Let $\Xi_{n,h}$ denote the denominator in (5.7) and let $\Xi_{n,h,\kappa}$ denote the denominator in (5.8). The same reasoning leading to (5.7) can be used to show that $T_{1,n} \xrightarrow{\mathcal{L}} N((\xi^h/\Xi_{n,h}), 1)$ under either H_0 or H_1. Recall from Section 5.2 that under $H_0 : \xi = 0$, it is clear that $\xi = \xi^h$.

Now let c_α denote a critical value from $N(0,1)$ for a test of size α such that for $\nu \sim N(0,1)$, $\Pr(\nu > c_\alpha) = 1 - \Phi(c_\alpha)$ (e.g., if $\alpha = 0.05$, then $c_\alpha \approx 1.96$). Then

$$\begin{aligned}\lim_{n\to\infty} \Pr(T_{1,n} > c_\alpha) &= \lim_{n\to\infty} \Pr\Big(\nu + \frac{\xi^h}{\Xi_{n,h}} > c_\alpha\Big) \\ &= 1 - \lim_{n\to\infty} \Phi\Big(c_\alpha - \frac{\xi^h}{\Xi_{n,h}}\Big) \\ &= 1 - \lim_{n\to\infty} \Phi\Big(c_\alpha - \frac{\xi + O(h)}{\Xi_{n,h}}\Big) \end{aligned} \tag{5.9}$$

because, by Assumption 4.4, $\xi - \xi^h = O(h)$ when $h \to 0$. Then for any $\xi > 0$ and any sequence $h \to 0$, this probability approaches 1 as $n \to \infty$ because $\Xi_{n,h} \to 0$ as $n \to \infty$ and therefore tests based on $T_{1,n}$ are consistent. Similar reasoning leads to $\lim_{n\to\infty} \Pr(T_{2,n} > c_\alpha) = 1 - \lim_{n\to\infty} \Phi(c_\alpha - (\xi + O(h))/\Xi_{n,h,\kappa})$, and hence tests based on $T_{2,n}$ are also consistent.

5.4. *Power against asymptotic local alternatives*

In Section 5.3, the power of the tests tends to 1 as $n \to \infty$ for any fixed $\xi > 0$. While this is perhaps not surprising, additional insight is gained by considering the power of the tests (based on either $T_{1,n}$ or $T_{2,n}$) under a sequence of local alternatives $H_1 : \xi_n = an^{-\rho}$ for some $a,\ \rho > 0$ and finding values of ρ such that the tests remain consistent.

Consider, first, the test based on $T_{1,n}$ (for appropriately small values of $p + q$). Because $n_1 = nc_1$ and $n_{2,h} = nh^r c_2$ for constants $c_1,\ c_2 > 0$ (typically, $c_1 = c_2 = 0.5$), it is easy to show that

$$\Xi_{n,h} = n_h^{-1/2}\sqrt{h^r(c_2/c_1)\widehat{\sigma}^2_{n_1} + \widehat{\sigma}^{2,h}_{c,n_2}},$$

where the square-root term converges in probability to σ_c as $n \to \infty$ and $h \to 0$. From this, it can be seen that for a sequence of alternatives with $\xi_n = an_n^{-1/2}n^\epsilon$ for some $\epsilon > 0$, the power in (5.9) still converges to 1 as $n \to \infty$. If the optimal order is used for h, then this gives the sequence $\xi_n = an^{-\rho}$ and the test remains consistent provided $\rho < (1/2(\kappa r + 1))$.

For larger values of $p + q$ when the test statistic $T_{2,n}$ must be used, similar reasoning provides

$$\Xi_{n,h,\kappa} = n_h^{-\kappa}\sqrt{h^{r\kappa}(c_2/c_1)\widehat{\sigma}^2_{n_1} + \widehat{\sigma}^{2,h}_{c,n_2}},$$

where again the square-root term converges in probability to σ_c when $n \to \infty$ and $h \to 0$. Unlike the previous case, here the sequence of alternatives must depend also on κ. Writing $\xi_n = ah_h^{-\kappa}n^\epsilon$ for some $\epsilon > 0$, it can be seen that using the optimal order for h provides the sequence of alternatives $\xi_n = an^{-\rho}$, and the test remains consistent provided $\rho < \kappa/(\kappa r + 1)$.

In either case (i.e., where either $T_{1,n}$ or $T_{2,n}$ must be used), the results here demonstrate that the asymptotic local power of the test depends on the dimensionality of the problem through both r and $p + q$. This is seen in the simulation results presented in online Appendix E.

5.5. *Bandwidth optimization*

As noted above, explicit expressions for the two components Δ_1 and Δ_2 of the estimation error in (4.8) are not available. Consequently, the best that can be done is to determine the order of optimal bandwidths by balancing the order of the two error terms yielding $h \propto n^{-1/(r+\kappa^{-1})}$ as explained earlier. Although the order by itself is of little help in applications, following the suggestion of Jeong et al. (2010) we can select optimal bandwidths for estimating the conditional distribution $H_{XY|Z}(x, y \mid z)$ by $\widehat{H}_{XY|Z}(x, y \mid z)$ given in (3.2). This can be accomplished using the least-squares cross-validation (LSCV) procedure described by Li et al. (2013), smoothing only on the r conditioning variables in Z, and not the dependent variables (X, Y). Note that, as proved by Hall et al. (2004), if one component of Z is irrelevant, then the corresponding bandwidth obtained by LSCV will converge to infinity as $n \to \infty$; but for relevant components of Z, LSCV gives a bandwidth with optimal rate $h \propto n^{-1/(r+4)}$ for estimating $H_{XY|Z}(x, y \mid z)$.

Recall that if Z is relevant, the optimal bandwidths for estimating $\lambda(x, y \mid z)$ have a different order ($h \propto n^{-1/(r+\kappa^{-1})}$, as opposed to $h \propto n^{-1/(r+4)}$) because of the presence of the noise due to localizing represented by Δ_2 in (4.8). In practice, it is possible to optimize bandwidths using LSCV, and then to correct the resulting bandwidths by multiplying by the scaling factor $n^{1/(r+4)} n^{-1/(r+\kappa^{-1})} = n^{(\kappa^{-1}-4)/((r+4)(r+\kappa^{-1})}$ to obtain bandwidths h with optimal order for estimating $\lambda(x, y \mid z)$. To avoid numerical difficulties, for the jth element Z_i^j of Z_i, $j = 1, \ldots, r$, $i = 1, \ldots, n$, one should, in practice, bound the LSCV search between a small factor (i.e., 0.01) times the normal reference rule bandwidth (i.e., $0.01 \times 1.06 \widehat{\sigma}_j n^{1/5}$, where $\widehat{\sigma}_j$ is the sample standard deviation of the observations Z_i^j, $j = 1, \ldots, n$) and two times the difference $(\max_i(Z_i^j) - \min_i(Z_i^j))$. If Z_i^j is irrelevant, LSCV will drive the jth element h_j of h to its upper bound; using a bounded kernel (e.g., the Epanechnikov kernel), no smoothing will be done in the jth dimension of Z when this happens. In such cases, there is no need to apply the scaling factor above to h_j.[10]

5.6. *Replicability*

It is important to note that tests based on the statistics defined in (5.7) and (5.8) are valid for any split of a given sample of size n into mutually exclusive, collectively exhaustive subsamples of sizes n_1 and n_2. However, there are $n!/((n_1!)(n_2!))$ possible splits (e.g., for $n = 100$ and $n_1 = \lfloor n_1 \rfloor$, $n_2 = n - n_2$ there are more than 10^{25} possible splits), and results can vary over these splits. This means that two researchers using the same data might reach different results by using different splits of the sample. Worse, a naïve or dishonest researcher might be tempted to split the sample repeatedly until the desired result is obtained.

It does not appear to be possible to combine information across many splits of a given sample and to obtain meaningful results. One might split the sample randomly (i.e., 100 or 1,000 times) and then average the resulting values of the test statistic from (5.7) or (5.8), but the values are not independent across the different sample splits, and the covariance is of complicated and unknown form.

In order to make results of our tests repeatable and verifiable, we propose a deterministic algorithm to randomly split n observations on $(p + q + r)$ variables. The rule is described and implemented in the R programming language in online Appendix D, where some examples illustrating usage are also presented.

5.7. *Performance of the tests*

Online Appendix E presents results from three sets of Monte Carlo experiments that provide evidence on the performance of the tests of separability proposed above for a variety of sample sizes and dimensionalities. The results of the experiments indicate that in most applied settings, one can reasonably expect the tests to give sizes close to nominal sizes, and with the power of the tests increasing with increasing departures from the null. The curse of dimensionality is present, of course, but as shown in the next section dimension-reduction methods can be used to help mitigate the effects of large dimensions.

[10] Theorem 2.2 of Li et al. (2013) establishes asymptotic equivalence of our data-driven bandwidths selected by cross-validation and the optimal, non-stochastic bandwidths. Consequently, the results described in Theorems 4.2–4.4 remain valid when optimal bandwidths are replaced with our cross-validated bandwidths.

6. EMPIRICAL ILLUSTRATION

As a final exercise, we revisit the empirical examples provided by Simar and Wilson (2007), where the estimated efficiency of US banks is regressed on some explanatory variables in a second-stage analysis. We start with the same data used by Simar and Wilson (2007), and we consider both the subsample of 322 banks as well as the full sample of 6,955 banks examined by Simar and Wilson (2007). The data include observations on three inputs (purchased funds, core deposits and labour) and four outputs (consumer loans, business loans, real estate loans and securities held). The data also include observations for two continuous explanatory variables used by Simar and Wilson (2007): *SIZE* (i.e., the log of total assets, reflecting the sizes of the banks) and *DIVERSE* (i.e., a measure of the diversity of banks' loan portfolios). Specific definitions of variables and other data details are given in Simar and Wilson (2007).

Our empirical examples here and in Simar and Wilson (2007) are motivated by Aly et al. (1990). They similarly estimate efficiency for a sample of 322 US banks operating during the fourth quarter of 1986. Then they attempt to explain variation in the first-stage efficiency estimates in a second-stage regression by regressing estimated efficiency on continuous variables reflecting bank size and loan-type diversity, as well as binary dummy variables reflecting membership in a multi-bank holding company and presence in a metropolitan statistical area. Whereas Aly et al. (1990) used the second-stage regression in an attempt to better understand the performance of US banks' operations, Simar and Wilson (2007) carefully note that their second-stage regressions are only for purposes of illustrating the bootstrap methods for inference developed in their paper. As discussed above, and as noted by Simar and Wilson (2007), such second-stage regressions can only be meaningful if the separability condition in Assumption 2.1 holds. Simar and Wilson (2007) also noted that this condition should be tested before employing a second-stage regression, but until now no such test has been available.

It is well known that the distribution of US bank sizes is heavily skewed to the right; in fact, the distribution of total assets of US banks is roughly log–log-normal; see, e.g., Wheelock and Wilson (2001) for a discussion. In order to use global bandwidths, as opposed to adaptive bandwidths (which would increase computational burden), we first eliminate very large banks and other outliers from the subsample of 322 observations as described by Florens et al. (2014) (who used the same data in an empirical illustration), leaving 303 observations for analysis. Similarly, we omit the largest 5% of banks from the full sample of 6,955 observations, leaving 6,607 observations. To further reduce computational burden, we exploit multicollinearity among the input and output variables by aggregating inputs into a single measure; we also aggregate outputs into a single measure using eigensystem techniques employed by Florens et al. (2014) in their analysis of the subsample of our data and as described by Daraio and Simar (2007a, pp. 148–50). Because of the high degrees of correlation among the original input and output variables, little information is lost by this aggregation, while dimensionality is reduced from $(p + q) = 7$ to 2.

We test the separability condition (Assumption 2.1) using both the subsample of 303 observations and the full sample of 6,607 observations using DEA estimators in both input and output directions, with bandwidths optimized by least-squares cross-validation and then adjusted to obtain the optimal order as discussed above. We first test separability marginally by considering only *SIZE*, and then by considering only *DIVERSE* so that $r = 1$. We also perform joint tests ($r = 2$) considering both *SIZE* and *DIVERSE*.

In all cases, we reject the null hypothesis (i.e., Assumption 2.1) in favour of the alternative hypothesis (i.e., Assumption 2.2) with p-values much less than 0.00005. In the individual tests where $r = 1$, we reject with *SIZE* more strongly than with *DIVERSE*. With the joint tests where $r = 2$, the values of the test statistics are between those where we test only with *SIZE* and only with *DIVERSE*, as one would expect.

The rejection of separability with respect to *SIZE* is hardly surprising given that larger banks necessarily can produce more output than smaller banks. Of course, *SIZE* is highly correlated with banks' inputs and outputs. Nonetheless, this variable is used by by Aly et al. (1990) in their second-stage regression, and one must assume separability in order to believe the second-stage estimation makes any sense at all. Moreover, Aly et al. (1990) are not the only ones to use such variables in second-stage regressions. The rejection with respect to *DIVERSE* is less obvious *a priori*, and suggests that conditional efficiency estimators should be used to analyse efficiency among banks.[11]

7. CONCLUSIONS

We have provided CLTs for conditional efficiency estimators, allowing researchers to estimate confidence intervals for mean conditional efficiency or to compare mean conditional efficiency across groups of producers analogous to the test of equivalent mean unconditional efficiency developed in Kneip et al. (2016). We have also provided a test of the restrictive separability condition described by Simar and Wilson (2007) on which many papers that regress estimated efficiency scores on some environmental variables depend. We prove consistency of the test and examine its local power. Now the assumption of separability can be tested empirically. In our empirical example in Section 6, patterned after the application by Aly et al. (1990), we easily reject separability suggesting that the results of their second-stage regression are meaningless, or at best very difficult to interpret. Furthermore, it raises the question of whether separability would similarly be rejected in the hundreds or thousands of papers that have regressed estimated efficiencies on environmental variables in a second-stage regression. It is perhaps too much to expect that all of these studies be re-examined, but now that an easily implemented test of separability has been made available, researchers should employ the test before proceeding to a second-stage regression. Moreover, whenever the test rejects separability, the researcher should use conditional efficiency estimators instead of unconditional estimators in order to estimate distance to the frontier of Ψ^z instead of the frontier of Ψ, which has no particular economic meaning when separability does not hold. Whenever separability is rejected, the new CLT results will be useful tools for empirical researchers.

Of course, failure to reject the null hypothesis of separability does not by itself imply that separability holds. As is always the case, our test can do only one of two things: it can either reject or fail to reject the null hypothesis. Failure to reject might be due to other factors, such as insufficient data, or too many dimensions. In the latter case, we show in the empirical example of Section 6 how dimensionality can be reduced before testing separability.

[11] Note that the second-stage regression in Simar and Wilson (2007) was used only to illustrate how one might apply the bootstrap methods proposed there. However, results from the second-stage regression in Aly et al. (1990), and those from similar exercises in other papers that have regressed estimates of bank efficiency on total assets, are rendered dubious and likely meaningless by the results obtained here.

It should be remembered, as noted in Section 3, that the conditional efficiency estimators provide consistent estimates regardless of whether separability holds, but the unconditional efficiency estimators provide meaningfully consistent estimates if and only if separability holds. Of course, if separability holds, the unconditional estimators converge faster than their conditional counterparts. However, when testing separability, these points argue in favour of a conservative test. Whereas one might ordinarily test a null hypothesis at the 10%, 5%, or 1% level, here one might want to test at a 20%, 30%, 40%, or even larger, level. The cost of a type-I error is slower convergence due to subsequent use of the conditional efficiency estimators, whereas the cost of a type-II error is loss of any statistical or economic meaning due to subsequent inappropriate use of unconditional efficiency estimators. The cost of a type-II error here is arguably greater than the cost of a type-I error, which is the reverse of the usual situation in hypothesis testing. Here, however, reversing things by testing a null hypothesis of non-separability versus an alternative hypothesis of separability would result in a test with poor size and power properties, as separability is a much more restrictive condition than non-separability.

ACKNOWLEDGEMENTS

We gratefully acknowledge research support from Sapienza Research Awards 2015 N. C26H15XNFS, IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) and from US National Science Foundation grant no. SMA-1243436. We have used the Palmetto Cluster maintained by Clemson Computing and Information Technology (CCIT) at Clemson University, and we are grateful for technical support from the staff of CCIT. Any remaining errors are solely our responsibility.

REFERENCES

Aly, H. Y., R. G. Grabowski, C. Pasurka and N. Rangan (1990). Technical, scale, and allocative efficiencies in U.S. banking: an empirical investigation. *Review of Economics and Statistics 72*, 211–18.

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*, 1078–92.

Bădin, L., C. Daraio and L. Simar (2012). How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research 223*, 818–33.

Bădin, L., C. Daraio and L. Simar (2017). Bandwidth selection issues for nonparametric conditional efficiency. Working Paper, Institut de Statistique Biostatistique et Sciences Actuarielles, Université Catholique de Louvain.

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

Charnes, A., W. W. Cooper and E. Rhodes (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research 2*, 429–44.

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

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

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

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*, 243–67. Amsterdam: North-Holland.

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

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society*, Series A *120*, 253–81.

Florens, J. P., L. Simar and I. Van Keilegom (2014). Frontier estimation in nonparametric location-scale models. *Journal of Econometrics 178*, 456–70.

Hall, P., J. S. Racine and Q. Li (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association 99*, 1015–26.

Jeong, S. O. and L. Simar (2006). Linearly interpolated FDH efficiency score for nonconvex frontiers. *Journal of Multivariate Analysis 97*, 2141–61.

Jeong, S. O., B. U. Park and L. Simar (2010). Nonparametric conditional efficiency measures: asymptotic properties. *Annals of Operations Research 173*, 105–22.

Kneip, A., L. Simar and P. W. Wilson (2008). Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models. *Econometric Theory 24*, 1663–97.

Kneip, A., L. Simar and P. W. Wilson (2011). A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators. *Computational Economics 38*, 483–515.

Kneip, A., L. Simar and P. W. Wilson (2015). When bias kills the variance: central limit theorems for DEA and FDH efficiency scores. *Econometric Theory 31*, 394–422.

Kneip, A., L. Simar and P. W. Wilson (2016). Testing hypotheses in nonparametric models of production. *Journal of Business and Economic Statistics 34*, 435–56.

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

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

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

Simar, L. and P. W. Wilson (2011). Two-stage DEA: caveat emptor. *Journal of Productivity Analysis 36*, 205–18.

Simar, L. and P. W. Wilson (2013). Estimation and inference in nonparametric frontier models: recent developments and perspectives. *Foundations and Trends in Econometrics 5*, 183–337.

Simar, L. and P. W. Wilson (2015). Statistical approaches for non-parametric frontier models: a guided tour. *International Statistical Review 83*, 77–110.

Simar, L., A. Vanhems and I. Van Keilegom (2016). Unobserved heterogeneity and endogeneity in nonparametric frontier estimation. *Journal of Econometrics 190*, 360–73.

Wheelock, D. C. and P. W. Wilson (2001). New evidence on returns to scale and product mix among U.S. commercial banks. *Journal of Monetary Economics 47*, 653–74.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Appendices
Replication files