

METHODOLOGIES FOR ASSESSING GOVERNMENT EFFICIENCY

Caitlin O'Loughlin, Léopold Simar and
Paul Wilson

LIDAM Discussion Paper ISBA/2021/02

METHODOLOGIES FOR ASSESSING GOVERNMENT EFFICIENCY

CAITLIN O’LOUGHLIN LÉOPOLD SIMAR PAUL W. WILSON*

December 2020

Abstract

Nonparametric methods are widely used for assessing the performance of firms and other organizations in the private and public sectors. Typically, FDH or DEA estimators that estimate the attainable sets and its efficient boundary by enveloping the cloud of observed units in the appropriate input-output space are used. The statistical properties of these estimators have been established and inference is available using appropriate nonparametric techniques. In particular, hypotheses on model structure and the production process can be tested using recent theoretical results including Central Limit Theorems on limiting distribution of means of efficiency scores. This chapter shows how these results can be used for testing the equality of means of efficiency, convexity of production sets and separability with respect to environmental factors are considered, and in addition for analyzing the dynamics of the production process over time. An empirical illustration is provided by using the various results and tests to examine the performance of municipal governments in the U.S. in providing local public goods.

*O’Loughlin: Federal Reserve Bank of Philadelphia, Department of Supervision, Regulation, and Credit, Philadelphia, Pennsylvania, USA; email: caitlin.oloughlin@phil.frb.org.

Simar: Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain-la-Neuve, Louvain-la-Neuve, Belgium; email: leopold.simar@uclouvain.be.

Wilson: Department of Economics and School of Computing, Clemson University, Clemson, South Carolina, USA; email: pww@clemson.edu.

We thank the Cyber Infrastructure Technology Integration group at Clemson University for technical support and operating the Palmetto Cluster. Any remaining errors are solely our responsibility.

1 Introduction

Microeconomic theory of the firm is concerned with the ability of firms to transform inputs such as labor, physical capital, energy and materials into outputs that might include various goods or services depending on a particular industry. Standard graduate-level textbook treatments (e.g., Henderson and Quandt, 1971; Varian, 1978; or Mas-Colell et al., 1995) posit cost-minimizing and profit-maximizing behavior among firms and typically assume firms do these things optimally, although some (e.g., Lesourne, 1972) consider public-sector institutions. But the theory of production is much larger. The theory of production and efficiency analysis examines how production units (which might be for-profit firms, non-profit organizations, government entities such as agencies or local governments, or other decision-making units such as hospitals, schools, etc.) transform their inputs into outputs, and in addition *how well* they do so. Efficiency analyses have been developed to examine the performance of not only for-profit firms, but also for decision-making units that might not explicitly have maximization as a goal.

Production theory (as opposed to microeconomic theory of the firm) builds on the pioneering work of Koopmans (1951) and Debreu (1951). Farrell (1957) and Afriat (1972) were among the first to look at the empirical issues of how to estimate production efficiency. The efficient production frontier is defined in an appropriate input-output space as the locus of optimal combinations of input and output quantities. One may consider either the locus of the maximal attainable level of outputs corresponding to a given level of inputs, or when output quantities are regarded as exogenous, the researcher may search for the minimal levels of inputs necessary for producing a given amount of outputs. We will see below that more general situations can be also analyzed.

In this chapter, we discuss and illustrate modern, state-of-the-art methods that can be used to analyze and quantify in particular the productive performance of local, municipal governments. Municipal governments provide varying bundles of goods, services and amenities for residents, who in turn are free to choose among municipalities and hence the varying offerings. In addition to police and fire protection, municipal governments may provide roads and streets, traffic management, trash collection, street cleaning, water services, libraries, and other services. In principle, municipalities compete with each other both in terms of taxation as well as provision of services along the lines of Tiebout (1956) and Ellickson (1971). Moreover, while municipal governments offer varying bundles of goods and services, they also vary in terms of *how well* or how efficiently they provide their particular bundles, i.e., in terms of levels of waste that are incurred.

A number of studies have examined the production performance of municipal governments. For example, Grosskopf et al. (2001) suggest that competition among municipalities may create incentives to provide services efficiently by influencing citizens' willingness to pay for public services or their inclination to remain in the jurisdiction. Resident satisfaction over time also likely plays a role in this decision; Tran and Dollery (2020) note a strong relationship with this and municipal efficiency across different categories of municipal governments. Hayes et al. (1998), Grossman et al. (1999) and others find empirical evidence that competition among local governments tends to enhance efficiency. At the same time, friction caused by real estate transaction fees, costs of commuting and job search and other factors may reduce competition among municipalities, perhaps leading to inefficient provision of services.

While there is a substantial literature on municipal efficiency, most studies offer little or no justification for the choice of the estimators used, and few if any provide statistical inference, presenting instead only point estimates with no accompanying measure or indication of the surrounding uncertainty. Examples can be found in many of the studies reviewed by Tang, 1997; De Borger and Kerstens, 2000; Afonso, 2008; Da Cruz and Marques, 2014; de Oliveira Junqueira, 2015; and Narbón-Perpiñá and De Witte, 2018). Empirical analyses of efficiency among municipal governments can be broadly divided into those that employ fully parametric methods along the lines of Aigner et al. (1977) and Meeusen and van den Broeck (1977) versus those that use fully nonparametric methods such as the data envelopment analysis (DEA) estimators proposed by Farrell (1957) and popularized by Banker et al. (1984) or free disposal hull (FDH) estimators proposed by Deprins et al. (1984). The focus in this chapter is on nonparametric methods, which are more often used to analyze municipal efficiency than parametric methods. Among studies using nonparametric methods, DEA estimators are used much more often than FDH estimators to analyze production by local governments.¹ As seen below, the choice between FDH and DEA estimators is not innocuous.

Before efficiency or any other measure of productive performance can be estimated, a well-specified model of the production process is needed. Specifically, a well-specified statistical model is needed, for without such a model, one cannot know what is estimated nor is inference possible. Only within the context of a properly specified model can the “efficiency scores” of individual local governments be characterized. As seen below, these efficiency scores

¹The review by Narbón-Perpiñá and De Witte (2018, Table A2) lists 97 empirical studies of local government efficiency. Sixty-six of these studies use nonparametric estimators, while only 31 use parametric methods. Among the 66 papers employing nonparametric estimators, 50 use DEA estimators, 14 use FDH estimators, and 2 use both.

will be based on the distance from a particular municipality's input-output combination to the efficient frontier in a specific direction in the input-output space. The nonparametric model presented below involves minimal assumptions, and relies only on mild assumptions suggested by economic theory and, except for some additional mild regularity conditions, no parametric restrictions are imposed on the stochastic components of the model. As opposed to parametric models, the Data Generating Process (DGP) that specifies how observed data are generated in the model used below relies on only a few flexible assumptions. We describe recent theoretical developments that provide tools for making inference about efficiency as well as features of the model, including whether the production set is convex, returns to scale, the role of environmental variables and evolution of the frontier over time, as well as differences in mean efficiency across groups of municipalities.

The remaining part of this chapter is organized along the following lines. In the next section we define the basic model, concepts and notation for the production process. We then introduce some popular nonparametric estimators of the frontier and of efficiency, including the estimation of the efficiency scores. We summarize the main statistical properties of these estimators as well as estimation of confidence intervals for efficiency of individual decision-making units and recent Central Limit Theorem (CLT) results developed for averages of efficiency scores. In Section 4 we show how the CLT results can be used to build test statistics for different hypotheses about model features and how these might change over time. Section 7 illustrates of how the methods discussed in Sections 2–6 can be used to analyze the performance of local, municipal governments in the U.S. Conclusions are given in Section 8 with suggestions for some extensions.

2 A Nonparametric Model of Production

To establish notation and basic ideas, consider the vector $x \in \mathbb{R}_+^p$ of input quantities used to produce the output quantities $y \in \mathbb{R}_+^q$. The *production set*, i.e., the set of feasible combinations of input and output quantities is given by

$$\Psi = \{(x, y) \mid x \text{ can produce } y\}. \quad (2.1)$$

Of particular interest is the *technology*, or efficient boundary of Ψ defined as

$$\Psi^\partial = \{(x, y) \in \Psi \mid (\gamma^{-1}x, \gamma y) \notin \Psi \text{ for all } \gamma > 1\}. \quad (2.2)$$

We assume the usual characteristics for Ψ coming from microeconomic theory of the firm; i.e., Ψ is closed and both inputs and outputs are freely disposable.² We rule out free lunches with the assumption that production of any non-zero output requires use of a non-zero quantity of at least one input. Assuming free disposability of inputs and outputs amounts to assuming that the frontier is weakly monotonic, i.e., non-decreasing in inputs.

The efficiency of a production plan (x, y) is measured by its distance to the efficient boundary. This measure depends on the chosen direction to the frontier. The Debreu-Farrell input efficiency measure

$$\theta(x, y \mid \Psi) = \inf\{\theta > 0 \mid (\theta x, y) \in \Psi\} \quad (2.3)$$

is the radial distance of (x, y) to its projection on Ψ^θ along the ray x . In other words, $\theta(x, y \mid \Psi)$ is the smallest factor by which input quantities can be proportionately reduced while maintaining production at output levels y . Alternatively, the Debreu-Farrell output efficiency measure

$$\lambda(x, y \mid \Psi) = \sup\{\lambda > 0 \mid (x, \lambda y) \in \Psi\} \quad (2.4)$$

gives the maximum feasible, radial expansion of the outputs to reach the efficient frontier. Efficiency can also be measured in other directions. For example, the hyperbolic measure

$$\gamma(x, y \mid \Psi) = \sup\{\gamma > 0 \mid (\gamma^{-1}x, \gamma y) \in \Psi\} \quad (2.5)$$

suggested by Färe et al. (1985) gives the simultaneous, feasible proportionate reduction of input quantities and proportionate, feasible increase of output quantities for a unit operating at $(x, y) \in \Psi$ to reach the efficient frontier.

In most situations Ψ , and hence $\theta(x, y \mid \Psi)$, $\lambda(x, y \mid \Psi)$ and $\gamma(x, y \mid \Psi)$, are unknown and not observable. In empirical applications, the task at hand is to estimate these from a sample $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$ of observations on input and output quantities of individual production units. To understand what can be learned from the observations in $\mathcal{X}_n$, we need a statistical model that specifies the DGP, i.e., a set of assumptions that characterize how the observations are generated. As noted in Section 1, we focus here on nonparametric models requiring a minimal set of regularity assumptions.

To define the DGP, we assume the n observations (X_i, Y_i) are identically, independently distributed (iid) realizations of a random variables (X, Y) having Ψ as support. In other words, all observations lie within the production set, Ψ . Hence the joint density $f_{XY}(x, y)$ of (X, Y) is assumed to be zero everywhere outside Ψ . In addition, we assume this joint density

²Free disposability means that For any $\tilde{x} \geq x$ and $\tilde{y} \leq y$, if $(x, y) \in \Psi$ then $(\tilde{x}, \tilde{y}) \in \Psi$.

is strictly positive along the frontier (i.e., $f_{XY}(x, y) > 0 \forall (x, y) \in \Psi^\partial$), and continuously differentiable close to Ψ^∂ in the interior of Ψ so that there is not a probability mass along the frontier. To establish properties of DEA and FDH estimators, the frontier must also be sufficiently smooth. Assumptions regarding convexity of Ψ or constant versus variable returns to scale may be made as appropriate. For technical details and precise statements of the various assumptions needed, see Kneip et al. (2008, 2015b).

3 Nonparametric Estimation and Inference

3.1 Nonparametric Efficiency Estimators

Nonparametric estimation of the efficiency measures given in (2.3)–(2.5) requires replacing the unknown Ψ in the definitions of the measures with a suitable nonparametric estimator of Ψ . The most popular nonparametric estimators of Ψ based on the sample $\mathcal{X}_n$ are the *envelopment* estimators that envelop the sample observations in a minimal (subject to some conditions) set spanning the observed points $\{(X_i, Y_i)\}_{i=1}^n$. The FDH estimator, introduced by Deprins et al. (1984) relies only on the assumption of free disposability of inputs and outputs, and thus is the least-restrictive, most-flexible of the envelopment estimators. The FDH estimator is given by

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

i.e., the union of the positive orthants in the inputs and the negative orthants in the outputs, whose origin coincides with the data points in $\mathcal{X}_n$. Since under the assumptions of the model outlined in Section 2 all observations (X_i, Y_i) must lie in Ψ , necessarily $\widehat{\Psi}_{\text{FDH},n} \subseteq \Psi$. In this sense, $\widehat{\Psi}_{\text{FDH},n}$ is a biased estimator of Ψ , as it lies in the interior of Ψ and the intersection of $\widehat{\Psi}_{\text{FDH},n}$ and the complement of Ψ is the null set.

Alternatively, if one is willing to assume Ψ is a convex set, then the variable returns to scale (VRS) version of the DEA estimator can be used to estimate Ψ . The VRS-DEA estimator of Ψ is the convex hull of $\widehat{\Psi}_{\text{FDH},n}$, i.e., the smallest convex set containing $\widehat{\Psi}_{\text{FDH},n}$, and can be written as

$$\widehat{\Psi}_{\text{VRS},n} = \{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq \sum_{i=1}^n \omega_i Y_i, x \geq \sum_{i=1}^n \omega_i X_i, \sum_{i=1}^n \omega_i = 1, \omega_i \geq 0, i = 1, \dots, n\}. \quad (3.2)$$

If, in addition to convexity of Ψ , one is willing to assume constant returns to scale (CRS), then the CRS version of the DEA estimator can be used to estimate Ψ . The CRS-DEA

estimator of Ψ is given by

$$\widehat{\Psi}_{\text{CRS},n} = \{(x, y) \in \mathbb{R}_+^{p+q} \mid y \leq \sum_{i=1}^n \omega_i Y_i, x \geq \sum_{i=1}^n \omega_i X_i, \omega_i \geq 0, i = 1, \dots, n\}. \quad (3.3)$$

This is the conical hull of $\widehat{\Psi}_{\text{VRS},n}$ with vertex at the origin. By construction, when Ψ is convex and CRS holds, $\widehat{\Psi}_{\text{FDH},n} \subseteq \widehat{\Psi}_{\text{VRS},n} \subseteq \widehat{\Psi}_{\text{CRS},n} \subseteq \Psi$. If Ψ is convex and CRS does not hold, then necessarily $\widehat{\Psi}_{\text{FDH},n} \subseteq \widehat{\Psi}_{\text{VRS},n} \subseteq \Psi$ and $\widehat{\Psi}_{\text{CRS},n} \cap \bar{\Psi} \neq \emptyset$ where $\bar{\Psi}$ denotes the complement of Ψ . Both the VRS-DEA and CRS-DEA estimators were discussed by Farrell (1957), but received little attention until they were revisited and popularized by Charnes et al. (1978) and Banker et al. (1984).

Estimators of the efficiency of a particular production plan (x, y) are obtained by replacing Ψ in (2.3), (2.4) or (2.5) with one of the estimators of Ψ described above. For example, the input-oriented FDH efficiency estimator is $\widehat{\theta}_{\text{FDH},n}(x, y) = \theta(x, y \mid \widehat{\Psi}_{\text{FDH},n})$. This is easily computed as

$$\widehat{\theta}_{\text{FDH},n}(x, y) = \min_{i \in D_{x,y}} \max_{j=1, \dots, p} \left(\frac{X_i^{(j)}}{x^{(j)}} \right), \quad (3.4)$$

where $D_{x,y} = \{i \mid (X_i, Y_i) \in \mathcal{X}_n, X_i \leq x, Y_i \geq y\}$ is the set of indices of observations in $\mathcal{X}_n$ dominating the point (x, y) and $a^{(j)}$ denotes the j th component of a vector a . The DEA-VRS is obtained as the solution to the linear program

$$\widehat{\theta}_{\text{VRS},n}(x, y) = \min_{\theta, \omega} \{ \theta \mid y \leq \sum_{i=1}^n \omega_i Y_i, \theta x \geq \sum_{i=1}^n \omega_i X_i, \sum_{i=1}^n \omega_i = 1; \omega_i \geq 0 \forall i = 1, \dots, n \}. \quad (3.5)$$

The CRS-DEA estimator $\widehat{\theta}_{\text{CRS},n}(x, y)$ is obtained by dropping the constraint $\sum_{i=1}^n \omega_i = 1$ in (3.5). For computational details and other variations, see the detailed review in Simar and Wilson (2013).

3.2 Estimation and Confidence Intervals for Individual Efficiencies

The statistical properties of the nonparametric estimators defined above have been derived in a number of papers, including Kneip et al. (1998), Park et al. (2000), Kneip et al. (2008, (2015b) and (2016)), Park et al. (2010), Wilson (2011) and Daouia et al. (2017). See also Simar and Wilson (2015) and the references therein for a comprehensive review. To summarize, for a fixed point of interest (x, y) we have under mild regularity assumptions (see above)

$$n^\kappa \left(\widehat{\theta}_{\bullet,n}(x, y) - \theta(x, y \mid \Psi) \right) \xrightarrow{\mathcal{L}} Q_\bullet(\eta_{x,y}) \quad (3.6)$$

as $n \rightarrow \infty$, where κ determines the rate of convergence and depends on the assumptions and the chosen estimator with “•” representing FDH, VRS or CRS, and $Q_{\bullet}$ is a non-degenerate distribution depending on unknown parameters $\eta_{x,y}$. Hence the various nonparametric estimators that have been discussed are statistically consistent, with known limiting distributions. Moreover, the rates of convergence are also known. The convergence rate n^{κ} determines the order of the error of estimation: we have $\left(\hat{\theta}_{\bullet,n}(x, y) - \theta(x, y \mid \Psi)\right) = O_p(n^{-\kappa})$. The error of estimation becomes smaller in probability terms as the sample size n increases. The larger is κ , the faster the error decreases as n increases. Similar results hold for the output-oriented and hyperbolic efficiency estimators. In particular, the rate of convergence n^{κ} nor the form of the limiting distribution $Q_{\bullet}(\eta_{x,y})$ depend on the direction from the fixed point of interest (x, y) to the frontier, although the constants in $\eta_{x,y}$ may differ.

For the FDH estimator, provided free-disposability of inputs and outputs holds, $\kappa = 1/(p + q)$. Under free disposability, convexity of Ψ , and non-CRS, the VRS-DEA estimator has the rate n^{κ} with $\kappa = 2/(p + q + 1)$. With the additional assumption of CRS, the CRS-DEA estimator (as well as the VRS-DEA estimator, as proved by Kneip et al., 2016) converges at rate n^{κ} with $\kappa = 2/(p + q)$. The convergence rates illustrate the price of using nonparametric estimators; whereas parametric estimators typically converge at rate $n^{1/2}$, here the convergence rate of our nonparametric estimators becomes slower as the number $(p + q)$ of inputs and outputs increases. This is typical for nonparametric estimators, and is called the *curse of dimensionality*.³ However, obtaining the typical parametric, root- n rate depends crucially on having the correct model specification, which may be unlikely. If a parametric model is misspecified, then convergence may be rather fast, but the parametric estimators converge to some unknown thing that is not a feature of the underlying, true model. Robinson (1988) refers to this as “root- n inconsistency.” Consistency is arguably the most fundamental, important property of an estimator; if an estimator is consistent, then additional data will improve the estimate in probabilistic terms. If an estimator is inconsistent, then additional data may not be informative.

Another complication for the nonparametric efficiency estimators arises from the fact that the limiting distributions $Q_{\bullet}(\eta_{x,y})$ in (3.6) and similar expressions for the other estimators depend on unknown parameters in $\eta_{x,y}$ that are difficult to estimate. In the case of the FDH estimator, $Q_{\text{FDH}}(\eta_{x,y})$ is a Weibull distribution as shown in Park et al. (2000), but for the VRS-DEA case (Kneip et al., 2008) and for the CRS-DEA case (Park et al., 2010), the existence of the limiting distributions are proven, with the appropriate rates given above,

³As seen below, methods exist for reducing the dimensionality of a given problem. In many cases, this substantially reduces estimation error.

but explicit analytic forms for the distributions are not available. However, existence of the limiting distributions permits proof of the validity of the bootstrap for making inference and allows derivation of valid asymptotic confidence intervals for $\theta(x, y \mid \Psi)$, $\lambda(x, y \mid \Psi)$ and $\gamma(x, y \mid \Psi)$ in all the cases considered above (i.e., using either FDH, VRS-DEA or CRS-DEA estimators as appropriate). Technical details are given in Jeong and Simar (2006), Kneip et al. (2008, 2011) and Simar and Wilson (2011a).

3.3 Inference about Mean Efficiency

For testing hypotheses involving mean efficiency or other model features, more than just the asymptotic results for an individual point (x, y) given above is needed. In most cases, as seen below in Section 4, tests will be based on test statistics involving sample means of efficiency estimators evaluated at each of the random points (X_i, Y_i) contained in the sample $\mathcal{X}_n$. Hereafter, we drop the notation “•” in $\hat{\theta}_{\bullet, n}$ used above to indicate which estimator is used. That which is derived below is applicable to any estimator (i.e., FDH, VRS-DEA or CRS-DEA) and any orientation (i.e., input, output or hyperbolic) with the appropriate rate of convergence involving κ as discussed above.

To understand the problem, consider the simplest problem where the researcher wishes to analyze average efficiency over some sector of activity. To do this, we investigate how the sample mean $\hat{\mu}_n = n^{-1} \sum_{i=1}^n \hat{\theta}_n(X_i, Y_i)$ is related to the population mean $\mu_\theta = \mathbb{E}(\theta(X, Y \mid \Psi))$. It is important to note that the sample mean $\hat{\mu}_n$ involves *estimators* of efficiency, as opposed to the true efficiencies. The pioneering work of Kneip et al. (2015b) makes clear that even this seemingly simple problem there is no easy solution due to the inherent bias of the estimators. This bias remains of the same order $n^{-\kappa}$ after taking the mean over n observations, and consequently kills the variance which is proved by Kneip et al. (2015b) to be of order n^{-1} .

The basic, key theorem in Kneip et al. (2015b) may be summarized as follows. Under mild regularity conditions, which differ slightly according the chosen estimator, we have (from Kneip et al., 2015b, Theorem 4.1)

- (i) $\hat{\mu}_n$ is a consistent estimator of μ_θ , with a bias term of order $Cn^{-\kappa}$;
- (ii) $\hat{\sigma}_{\theta, n}^2 = n^{-1} \sum_{i=1}^n [\hat{\theta}_n(X_i, Y_i) - \hat{\mu}_n]^2$ is a consistent estimator of $\sigma_\theta^2 = \text{Var}(\theta(X, Y \mid \Psi))$; and
- (iii) the following basic limit theorem

$$\sqrt{n}(\hat{\mu}_n - \mu_\theta - Cn^{-\kappa} - R_{n, \kappa}) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2), \text{ with } R_{n, \kappa} = o(n^{-\kappa}) \quad (3.7)$$

as $n \rightarrow \infty$ holds under appropriate regularity conditions.

From these results, it is easy to see why problems arise in making inference about mean efficiency. If $\kappa > 1/2$, there is no problem and a simple Central Limit Theorem (CLT) derived from (3.7) is available, because $\sqrt{n}(Cn^{-\kappa} + R_{n,\kappa}) = o(1)$. However, $\kappa > 1/2$ if and only if $p + q \leq 1$ when FDH estimators are used, or $p + q \leq 2$ for the VRS-DEA case or $p + q \leq 3$ for the CRS-DEA case. If $\kappa = 1/2$ then the bias is a constant, and if $\kappa < 1/2$ the bias explodes. In either of these cases (i.e., whenever $\kappa \leq 1/2$), standard CLT results (e.g., the Lindeberg-Feller CLT) cannot be used. This means that when FDH estimators are used, the usual CLT results are *never* applicable. Moreover, replacing the scale factor $\sqrt{n}$ with n^ξ with $\xi < \kappa \leq 1/2$ is not a viable option, since this would cause the variance to converge to zero whenever $\kappa \leq 1/2$, rendering inference impossible.

For cases where $\kappa \leq 1/2$ (which is typical in applied work), the solution suggested by Kneip et al. (2015b) is two-fold. First, one must correct for the leading term $Cn^{-\kappa}$ in the inherent bias in (3.7) using a generalized jackknife estimator of the bias. Second, when necessary depending on the estimator and on the dimension $p + q$, one must compute the average on a subsample of the n efficiency estimates to control the variance of the resulting average.

The bias correction works along the following lines. As shown in Kneip et al. (2015b), a consistent estimator of the bias term $Cn^{-\kappa}$ is given by

$$\hat{B}_{n,\kappa} = (2^\kappa - 1)^{-1}(\bar{\mu}_{n/2}^* - \hat{\mu}_n), \quad (3.8)$$

where $\bar{\mu}_{n/2}^*$ is a jackknife analog of $\hat{\mu}_n$ obtained as

$$\bar{\mu}_{n/2}^* = (\bar{\mu}_{n/2}^{(1)} + \bar{\mu}_{n/2}^{(2)})/2, \quad (3.9)$$

where $\bar{\mu}_{n/2}^{(1)}$ is a version of $\hat{\mu}_n$ based on a random subset of $\mathcal{X}_n$ of size $n/2$ and $\bar{\mu}_{n/2}^{(2)}$ is the analog based on the remaining part of the sample (for simplicity, assume that n is even). Formally, one computes for $j = 1, 2$,

$$\bar{\mu}_{n/2}^{(j)} = 2n^{-1} \sum_{\{i | (X_i, Y_i) \in \mathcal{X}_{n/2}^{(j)}\}} \hat{\theta}_{n/2}(X_i, Y_i | \mathcal{X}_{n/2}^{(j)}), \quad (3.10)$$

where $\mathcal{X}_{n/2}^{(1)} \cap \mathcal{X}_{n/2}^{(2)} = \emptyset$ and $\mathcal{X}_{n/2}^{(1)} \cup \mathcal{X}_{n/2}^{(2)} = \mathcal{X}_n$. Note carefully that it is explicit in the notation that the efficiency estimators (either FDH, VRS-DEA or CRS-DEA) are computed using the limited reference samples $\mathcal{X}_{n/2}^{(j)}$. The intuition behind (3.8) is that the $\bar{\mu}_{n/2}^{(j)}$, $j = 1, 2$, are two estimators of μ_θ each having the same asymptotic bias as $\hat{\mu}_n$, but with n replaced by $n/2$.

Kneip et al. (2015b) prove that $\widehat{B}_{n,\kappa} = Cn^{-\kappa} + o(n^{-\kappa}) + o_p(n^{-1/2})$. Kneip et al. (2016) suggest repeating the exercise just described a large number of times (say, L times) providing L estimates of the bias, and then averaging the L estimates to reduce noise. Then for $\ell = 1, \dots, L$, we have, based on the ℓ th random share of $\mathcal{X}_n$ in two parts, as above, the estimates $\widehat{B}_{n,\kappa}^\ell$ given in (3.8). These are then used in

$$\widehat{B}_{n,\kappa} = L^{-1} \sum_{\ell=1}^L \widehat{B}_{n,\kappa}^\ell \quad (3.11)$$

to estimate the required bias correction. Averaging over the L bias estimates reduces the variance of the final bias estimate (in practice, selecting $L = 20$ is typically sufficient).

The bias correction solves the main part of the bias problem, but the remainder $R_{n,\kappa} = o(n^{-\kappa})$ must also be controlled. This is achieved, when $p + q$ is too large, by limiting the number of efficiency estimates used to compute $\widehat{\mu}_n$. When $\kappa \leq 1/2$, a subsample version of $\widehat{\mu}_n$ given by

$$\widehat{\mu}_{n_\kappa} = n_\kappa^{-1} \sum_{\{j | (X_j, Y_j) \in \mathcal{X}_{n_\kappa}^*\}} \widehat{\theta}_n(X_j, Y_j | \mathcal{X}_n), \quad (3.12)$$

must be used, where the average is taken over a random subsample $\mathcal{X}_{n_\kappa}^* \subseteq \mathcal{X}_n$ of size $n_\kappa = \lfloor n^{2\kappa} \rfloor \leq n$, where $\lfloor a \rfloor$ denotes the largest integer not greater than $a \in \mathbb{R}$. However, the individual efficiency estimates are computed using the full sample as the reference set, as seen explicitly in the notation. Due to the particular expressions of $R_{n,\kappa}$ which depend on the chosen estimator, Kneip et al. (2015b) derive (under the same regularity assumptions as above) the following CLTs.

- (i) For $p + q \leq 3$, if the FDH estimator is used, or for $p + q \leq 4$ and if Ψ is convex and the DEA-VRS estimator is used, or for $p + q \leq 5$ and if Ψ is convex and Ψ^∂ exhibits CRS and either the CRS-DEA or the VRS-DEA estimator is used, then

$$\sqrt{n} \left(\widehat{\mu}_n - \mu_\theta - \widehat{B}_{n,\kappa} - R_{n,\kappa} \right) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2), \text{ with } R_{n,\kappa} = o(n^{-1/2}) \quad (3.13)$$

as $n \rightarrow \infty$; and

- (ii) when $\kappa < 1/2$, with κ depending on the chosen estimator as discussed above, then

$$\sqrt{n^\kappa} \left(\widehat{\mu}_{n_\kappa} - \mu_\theta - \widehat{B}_{n,\kappa} - R_{n,\kappa} \right) \xrightarrow{\mathcal{L}} N(0, \sigma_\theta^2), \text{ with } R_{n,\kappa} = o(n^{-\kappa}) \quad (3.14)$$

as $n \rightarrow \infty$.

Note that either of the two CLTs in (3.13) and (3.14) can be used with the FDH estimator when $p + q = 3$ (or when $p + q = 4$ with VRS-DEA under convexity, or when $p + q = 5$ under convexity and CRS, using either the VRS-DEA or CRS-DEA estimator). But as explained in Kneip et al. (2015b), in these particular cases, the second result in (3.14) provides better approximations in finite samples.

The CLTs (3.7), (3.13) and (3.14) provide tools for building asymptotic confidence intervals for μ_θ . For example, suppose $p + q \geq 4$, Ψ is assumed convex, and VRS-DEA estimators are used. Using (3.14), an asymptotically correct $(1 - \alpha)$ confidence interval for μ_θ is given by

$$\left[\hat{\mu}_{n_\kappa} - \hat{B}_{n,\kappa} \pm z_{1-\alpha/2} \hat{\sigma}_{\theta,n} / n^\kappa \right], \quad (3.15)$$

with $\kappa = 2/(p + q + 1)$ and where $z_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard normal distribution. These basic CLTs also provide tools that are useful for solving various testing problems as described below in Sections 4–6. Each test described in Sections 4–6 involves comparison of two sample means of efficiency estimates.

3.4 Dimension Reduction

The curse of dimensionality discussed above in Section 3.2 may jeopardize the quality and the usefulness of both estimates and any statistical inference in nonparametric frontier models if the dimension of the problem, $p + q$, is “too large” relative to the given sample size n . Wilson (2018) discusses a number of “rules-of-thumb” for deciding whether dimensionality is “too large” that have appeared in the literature, and also discusses why these rules are not useful. In many applied problems, substantial multicollinearity exists among the outputs and among the inputs, and this can often be exploited to reduce the dimensionality of the problem. Daraio and Simar (2007) suggest a method for reducing dimensionality based on factor analysis and exploiting near-collinearity among inputs or outputs when it exists. Wilson (2018) provides several diagnostics for assessing whether the number of dimensions in a given problem is excessive for the available sample size, and for indicating whether dimension-reduction is likely to reduce estimation error. Wilson (2018) also provides extensive Monte Carlo results to quantify the tradeoff between information lost versus the reduction in estimation error (in terms of mean-square error) due to dimension reduction.

The method developed by Daraio and Simar (2007) is based on the singular value decomposition of a moment matrix. Dimension-reduction may be applied to either inputs, outputs, or both. Here, we illustrate the idea in terms of inputs. Consider the case where some inputs are closely related. Let $\mathbf{X}$ denote the $(n \times k)$ matrix of observations on these inputs,

where $k \leq p$. The goal is to reduce dimensionality by transforming these k input variables to a single measure. Suppose the k columns of $\mathbf{X}$ have been standardized by dividing each column by its standard deviation. Then the $(k \times k)$ (raw) moment matrix is $\mathbf{X}'\mathbf{X}$. Let $\xi_1, \dots, \xi_k$ denote the eigenvalues of $\mathbf{X}'\mathbf{X}$ in decreasing order, and let $a_1, \dots, a_k$ denote the corresponding eigenvectors.

All of the data in $\mathbf{X}$ lie in the positive orthant $\mathbb{R}_+^k$. It is well-known in the statistical literature (e.g., see Härdle and Simar, 2019), that the best linear representation of the data in $\mathbf{X}$ in one dimension is given by the first “input factor” $F_x = \mathbf{X}a_1$, and the quality of the representation is measured by its percent of inertia given by $R_x = \xi_1 / \sum_{j=1}^k \xi_j$. The latter provides a measure of how close the original data observations in $\mathbf{X}$, in the k -dimensional space, are “near” their projections F_x on the first eigenvector. In other words, the ratio R_x measures the percentage of the information contained in the k columns of $\mathbf{X}$ that is shared by the first factor F_x . For instance, if $R_x = 0.95$, then F_x contains 95 percent of the information contained in $\mathbf{X}$. So for large values of R_x (which happens when the columns of $\mathbf{X}$ exhibit substantial collinearity), one might reasonably replace the k original inputs by this single univariate factor. As noted in Daraio and Simar (2007), the correlations of the original inputs, that have been aggregated, with this input factor indicates how well this new one-dimensional variable represents the original ones. Wilson (2018) uses Monte-Carlo experiments to provide guidelines based on R_x and sample size n for determining whether dimension reduction as described here is likely to reduce mean-square error of the efficiency estimates.

Wilson (2018) notes that some (e.g., Adler and Golany, 2007) reduce dimensionality by decomposing the correlation matrix for the columns of $\mathbf{X}$ instead of the moment matrix $\mathbf{X}'\mathbf{X}$. But for radial efficiency estimated by FDH, VRS-DEA or CRS-DEA estimators, the moment matrix is more appropriate, since it provides information on how tightly packed are rays from the origin to each observation and it is along these rays that efficiency is measured. By contrast, the correlation matrix involves central moments, and describes how the data vary around means of the input (or output) variables, which is irrelevant for purposes of radial efficiency estimates.

4 Testing Hypotheses about Model Features

4.1 Testing Equality of Mean Efficiency for Two Groups of Units

We begin with a relatively easy test. Suppose there are two independent samples of size n_1 and n_2 of firms belonging to two groups labeled G_1 and G_2 . We want to test whether $\mu_{1,\theta} = \mathbb{E}_{G_1}(\theta(X, Y) \mid \Psi_1)$ and $\mu_{2,\theta} = \mathbb{E}_{G_2}(\theta(X, Y) \mid \Psi_2)$ are equal against the alternative hypothesis that group 1 has a larger mean efficiency than group 2. Alternatively, one might test whether group 2 has larger mean efficiency than group 1, or one could simply test whether the two groups have different mean efficiency. The latter amounts to a two-sided test, while the two previous tests are one-sided tests. The test is developed by Kneip et al. (2016). Here, we present the most general version of the test, where no restriction are made on whether the attainable sets for the two groups are the same, i.e., whether units in the two groups face the same frontier. Kneip et al. (2016, Section 3.1.2) explain how the test can be adapted to situations where firms in the two groups face the same frontier.

The researcher is confronted with two iid samples, $\mathcal{X}_{1,n_1} = \{(X_i, Y_i)\}_{i=1}^{n_1}$ and $\mathcal{X}_{2,n_2} = \{(X_i, Y_i)\}_{i=1}^{n_2}$. Assume the two samples are independent of each other. The CLTs (3.13) or (3.14), depending on the dimension $p + q$ and the chosen estimator (FDH, VRS-DEA or CRS-DEA)) can be applied independently for each of the two groups and the corresponding sample means of estimated efficiencies in each group. For example, if $\kappa < 1/2$, then (3.14) applies and we have for $\ell = 1, 2$, as $n_\ell \rightarrow \infty$,

$$\sqrt{n_{\ell,\kappa}} \left(\hat{\mu}_{n_{\ell,\kappa}} - \mu_{\ell,\theta} - \hat{B}_{\ell,n_{\ell,\kappa}} - R_{n_{\ell,\kappa}} \right) \xrightarrow{\mathcal{L}} N(0, \sigma_{\ell,\theta}^2), \text{ with } R_{n_{\ell,\kappa}} = o(n_{\ell}^{-\kappa}), \quad (4.1)$$

where $n_{\ell,\kappa} = \lfloor n_\ell^{2\kappa} \rfloor \leq n_\ell$. Necessarily, two sequences of independent random variables each having a normal limiting distribution, possess a joint limiting bivariate normal distribution with zero covariance. Consequently, the difference of the two random independent sequences has a limiting normal distribution.

Using the CLTs in (3.13) and (3.14) developed by Kneip et al. (2015b) leads to test results for two different cases. First, if $p + q \leq 3$ when FDH estimators are used, or if $p + q \leq 4$, Ψ_1 and Ψ_2 are convex and the VRS-DEA estimator is used, or if $p + q \leq 5$, Ψ_1 and Ψ_2 are convex and the frontiers of both sets exhibit CRS and either the CRS-DEA or VRS-DEA estimators are used, then

$$\hat{\tau}_{1,n_1,n_2} = \frac{(\hat{\mu}_{1,n_1} - \hat{\mu}_{2,n_2}) - (\hat{B}_{1,n_1,\kappa} - \hat{B}_{2,n_2,\kappa}) - (\mu_{1,\theta} - \mu_{2,\theta})}{\sqrt{\hat{\sigma}_{1,\theta,n_1}^2/n_1 + \hat{\sigma}_{2,\theta,n_2}^2/n_2}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (4.2)$$

provided $n_1/n_2 \rightarrow c$ as $n_1, n_2 \rightarrow \infty$ where c is a constant. In the second case, where $\kappa < 1/2$ with κ depending on $p + q$ and the specific estimator used, we have

$$\widehat{\tau}_{2,n_1,\kappa,n_2,\kappa} = \frac{(\widehat{\mu}_{1,n_1,\kappa} - \widehat{\mu}_{2,n_2,\kappa}) - (\widehat{B}_{1,n_1,\kappa} - \widehat{B}_{2,n_2,\kappa}) - (\mu_{1,\theta} - \mu_{2,\theta})}{\sqrt{\widehat{\sigma}_{1,\theta,n_1}^2/n_{1,\kappa} + \widehat{\sigma}_{2,\theta,n_2}^2/n_{2,\kappa}}} \xrightarrow{\mathcal{L}} N(0, 1), \quad (4.3)$$

again provided $n_1/n_2 \rightarrow c$, as $n_1, n_2 \rightarrow \infty$ where c is a constant and for $\ell = 1, 2$, $n_{\ell,\kappa} = \lfloor n_\ell^{2\kappa} \rfloor \leq n_\ell$ and κ characterizes the rate of convergence in each group and has been defined above.

The results in (4.2) and (4.3) are valid for any values of $\mu_{1,\theta}$ and can be used to build asymptotic confidence intervals for their differences. To test e.g. $H_0: \mu_{1,\theta} = \mu_{2,\theta}$ against the alternative $H_1: \mu_{1,\theta} > \mu_{2,\theta}$ one can compute either τ_{1,n_1,n_2} or $\tau_{2,n_1,\kappa,n_2,\kappa}$, as appropriate, while setting $(\mu_{1,\theta} - \mu_{2,\theta}) = 0$ (i.e., under the null). Then for a test of size α , H_0 is rejected if the test statistic is larger than $\Phi^{-1}(1 - \alpha)$ where $\Phi^{-1}(\cdot)$ is the standard normal quantile function. Alternatively, the null can be rejected in favor of $H'_1: \mu_{1,\theta} < \mu_{2,\theta}$ when the test statistic is less than $\Phi^{-1}(\alpha)$. For a two-sided test, the null is rejected whenever the test statistic is less than $\Phi^{-1}(\alpha/2)$ or greater than $\Phi^{-1}(1 - \alpha/2)$ for a test of size α . Note that the independence of the two samples $\mathcal{X}_{1,n_1}$ and $\mathcal{X}_{2,n_2}$ is crucial, and avoids complications due to covariances.

Intensive Monte-Carlo experiments (see Table 1 in Kneip et al., 2016) indicate the test works reasonably well in practice with finite samples. As expected, the performances of the test (i.e., the achieved level and power) increases when the sample sizes increase for all the dimensions $(p + q)$ considered. Of course, for a fixed sample size, the performances decrease when the dimension increases: this advocates in practical problem with real data and suggests use of, if possible, the dimension reduction mentioned above.

4.2 Testing Convexity

Testing convexity is based on comparison of a sample mean of FDH efficiency estimates (in either the input, output or hyperbolic direction) which do not impose convexity, and a sample mean of VRS-DEA efficiency estimates (in the same direction) which impose convexity. The problem here is more complicated than in the preceding case, because we need the two sample means to be independent, and this requires, as suggested by Kneip et al. (2016), a random split of the original sample $\mathcal{X}_n$ into two mutually exclusive, collectively exhaustive subsamples. Kneip et al. (2016) establish CLTs based on this simple splitting device. The procedure described by Simar and Wilson (2020a) allows one to remove the ambiguity re-

sulting from a single random split by repeating the sample-slitting a large number of time in order to “integrates out” the randomness of a single split.

To formalize things, we wish to test the null hypothesis $H_1: \Psi$ is not convex. The assumption of free disposability of inputs and the outputs is maintained. Under the null, both the FDH and the VRS-DEA estimators are consistent, but under the alternative, only the FDH estimator is consistent. The researcher is faced with a single iid sample $\mathcal{X}_n$. By construction, $\hat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_n) \leq \hat{\theta}_{\text{FDH}}(X_i, Y_i \mid \mathcal{X}_n) \leq 1$, where again we make explicit in the notation the reference sample used to compute the estimates. Therefore $\hat{\mu}_{\text{FDH},n} - \hat{\mu}_{\text{VRS},n} \geq 0$. Under the null we expect $\hat{\mu}_{\text{FDH},n} - \hat{\mu}_{\text{VRS},n}$ to be “small,” while under the alternative the difference is expected to be “large.”

Using test statistics based on the difference between two means over all n observations in the sample cannot work, as explained by Kneip et al. (2016). If the null is true then $n^b (\hat{\mu}_{\text{FDH},n} - \hat{\mu}_{\text{VRS},n})$ converges to a degenerate distribution for any power $b \leq 1/2$ of n . Avoiding this problem requires randomly splitting the original sample into two independent subsamples $\mathcal{X}_{1,n_1}$ and $\mathcal{X}_{2,n_2}$. The sample mean of the VRS-DEA estimator evaluated at each $(X_i, Y_i) \in \mathcal{X}_{1,n_1}$ using only observations in $\mathcal{X}_{1,n_1}$ as the reference set is computed, and the sample mean of the FDH estimators evaluated at each $(X_i, Y_i) \in \mathcal{X}_{2,n_2}$ using only observations in $\mathcal{X}_{2,n_2}$ as the reference set is also computed. Note that here the two efficiency estimators have different rates of convergence, i.e., $\kappa_1 = 2/(p+q+1)$ for the VRS-DEA case and $\kappa_2 = 1/(p+q)$ for the FDH case. Kneip et al. (2016) suggest choosing n_1 and n_2 such that $n_1^{\kappa_1} = \lfloor n_2^{\kappa_2} \rfloor$ and $n_1 + n_2 = 2$.

For the VRS-DEA estimator, we have the analog of the quantities defined in Section 3.3 but with reference sample $\mathcal{X}_{1,n_1}$. This leads to the sample mean $\hat{\mu}_{\text{VRS},n_1}$, the sample variance $\hat{\sigma}_{\text{VRS},n_1}^2$ and the bias correction $\hat{B}_{\text{VRS},\kappa_1,n_1}$ (for explicit expressions, see Kneip et al., 2016). In situations where $\kappa_1 < 1/2$, the subsampling means

$$\hat{\mu}_{\text{VRS},n_1,\kappa_1} = n_{1,\kappa_1}^{-1} \sum_{\{(X_i, Y_i) \in \mathcal{X}_{1,n_1,\kappa_1}^*\}} \hat{\theta}_{\text{VRS}}(X_i, Y_i \mid \mathcal{X}_{1,n_1}) \quad (4.4)$$

are needed, where $n_{1,\kappa_1} = \lfloor n_1^{2\kappa_1} \rfloor < n_1$ for $\kappa_1 < 1/2$ and $\mathcal{X}_{1,n_1,\kappa_1}^*$ is a random subset of $\mathcal{X}_{1,n_1}$ of size n_{1,κ_1} . Note the summation in (4.4) is over the subsample, but the estimator under the summation sign is computed using the full reference set $\mathcal{X}_{1,n_1}$. Similarly, and without giving all the detailed expressions, using the subsample $\mathcal{X}_{2,n_2}$, we have for the FDH estimator the corresponding sample mean $\hat{\mu}_{\text{FDH},n_2}$, sample variance $\hat{\sigma}_{\text{FDH},n_2}^2$ and the bias

correction $\widehat{B}_{FDH,\kappa_2,n_2}$. If $\kappa_2 < 1/2$ the subsample version of the mean given by

$$\widehat{\mu}_{FDH,n_2,\kappa_2} = n_{2,\kappa_2}^{-1} \sum_{\{(X_i, Y_i) \in \mathcal{X}_{2,n_2,\kappa_2}^*\}} \widehat{\theta}_{FDH}(X_i, Y_i \mid \mathcal{X}_{2,n_2}) \quad (4.5)$$

is required, where the estimates under the summation sign are computed using all the observations in $\mathcal{X}_{2,n_2}$, but the summation is over only a subset of these observations, analogous to the VRS-DEA case in (4.4).

Depending on the value of $p + q$, different versions of the test statistic are available, since for each mean in the two independent groups, we have the CLTs described in (3.13) and (3.14). For $p + q \leq 3$ we have

$$\widehat{\tau}_{3,n} = \frac{(\widehat{\mu}_{FDH,n_2} - \widehat{\mu}_{VRS,n_1}) - (\widehat{B}_{FDH,n_2,\kappa_2} - \widehat{B}_{VRS,n_1,\kappa_1})}{\sqrt{\widehat{\sigma}_{FDH,\theta,n_2}^2/n_2 + \widehat{\sigma}_{VRS,\theta,n_1}^2/n_1}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (4.6)$$

as $n \rightarrow \infty$ under the null hypothesis of convexity. For $p + q > 3$, we have

$$\widehat{\tau}_{4,n} = \frac{(\widehat{\mu}_{FDH,n_2,\kappa_2} - \widehat{\mu}_{VRS,n_1,\kappa_1}) - (\widehat{B}_{FDH,n_2,\kappa_2} - \widehat{B}_{VRS,n_1,\kappa_1})}{\sqrt{\widehat{\sigma}_{FDH,\theta,n_2}^2/n_{2,\kappa_2} + \widehat{\sigma}_{VRS,\theta,n_1}^2/n_{1,\kappa_1}}} \xrightarrow{\mathcal{L}} N(0, 1) \quad (4.7)$$

as $n \rightarrow \infty$ under the null. The null hypothesis of convexity is rejected if the p -value $\widehat{p}_\ell = 1 - \Phi(\widehat{\tau}_{\ell,n})$ is suitable small (i.e., less than .1, .05 or .01) with $\ell \in \{3, 4\}$ indexing the appropriate statistic (depending on $p + q$) in (4.6) or (4.7).

Splitting the original sample $\mathcal{X}_n$ into two independent subsamples $\mathcal{X}_{1,n_1}$ and $\mathcal{X}_{2,n_2}$ is crucial for obtaining the non-degenerate limiting distributions of the test statistics. Simar and Wilson (2020a) show how a bootstrap can be used to remove the ambiguity resulting from a single random split. The idea is to repeatedly split the sample s times, and compute either (i) the sample mean of the appropriate test statistic in (4.6) or (4.7) computed for each of the s splits, or (ii) the one-sample Kolmogorov-Smirnov statistic comparing the empirical distribution of the s p -values corresponding to the test-statistics computed for each of the s sample-splits and the uniform distribution over $[0, 1]$. Due to the dependence between the s sample-splits, the distributions of these two test statistics are unknown. However, the bootstrap method provided by Simar and Wilson (2020a) permits easy estimation of p -values corresponding to the two statistics obtained from s sample-splits.

Results from Monte-Carlo experiments provided by Simar and Wilson (2020a) indicate that the procedure works well and achieves better performance than the method based on a single split, in terms of both size and power. The Kolmogorov-Smirnov statistic is also

observed to have slightly better properties than the averaged statistic. In practice, for a single sample of real data, the computational burden of the proposed bootstrap is feasible for modern desktop machines with sample sizes typically encountered in applied work.

The test of convexity of Ψ is extended by Kneip et al. (2016) and Simar and Wilson (2020a) for testing CRS against the alternative hypothesis of VRS (under the assumption of convexity for Ψ). The idea is analogous to the idea underlying the convexity test discussed here, and involves comparing sample means of DEA-VRS and DEA-CRS estimators. The issues are also similar; independence between the two sample means is required, which in turn requires randomly splitting the original sample into two independent subsamples. See Kneip et al. (2016) and Simar and Wilson (2020a) for specific details. All of the testing methods discussed so far have been implemented in the current version of the *R* package FEAR introduced by Wilson (2008).

5 Dealing with Environmental Factors

5.1 Conditional Frontiers

In applied work, there are often additional variables representing environmental factors that are neither inputs nor outputs in the usual sense, but which may affect production of outputs. When these environmental factors differ across producers, they should be taken into account, as failure to do so may lead to misleading results. Examples of environmental variables faced by local governments might include unobserved heterogeneity across regions within a country, differing legal and regulatory environments across states or provinces, differing forms of government across municipalities (e.g., mayors versus city managers) and other factors. These factors are beyond the control of the municipality, but may affect municipalities' ability to transform inputs into outputs in either of two ways. The environmental factors may (i) influence the shape of the attainable frontier Ψ^∂ , or (ii) influence the probability of being close to the efficient boundary Ψ^∂ by affecting the shape of the distribution of efficiency, or (iii) influence both Ψ^∂ and the distribution of efficiency.

Let Z denote a random vector of r environmental variables, with $Z \in \mathcal{Z} \subseteq \mathbb{R}^r$. The concept of *conditional frontiers* introduced by Cazals et al. (2002) and extended by Daraio and Simar (2005) provides a natural and appealing way to incorporate information contained in Z .⁴ The *conditional* attainable set is

$$\Psi^z = \{(x, y) \mid x \text{ can produce } y \text{ when } Z = z\} \quad (5.1)$$

⁴See also the recent surveys by Bădin et al. (2012, 2014).

for given levels $Z = z$ of the environmental factors. The *marginal* attainable set is then

$$\Psi = \{(x, y) \mid x \text{ can produce } y\} = \bigcup_{z \in \mathcal{Z}} \Psi^z. \quad (5.2)$$

In general, $\Psi^z \subseteq \Psi$, which includes the special case

$$\Psi^z = \Psi \text{ for all } z \in \mathcal{Z}. \quad (5.3)$$

This is known as the “separability” condition as described by Simar and Wilson (2007, 2011b). The separability condition in (5.3) is quite restrictive; it is one case among an uncountable set of possibilities. Simar and Wilson (2007) caution against merely assuming that the separability condition holds, and note that (5.3) should be tested against the alternative hypothesis that $\Psi^z \neq \Psi$ for some $z \in \mathcal{Z}$. However, as explained clearly by Simar and Wilson (2007, 2011b), (5.3) must hold in order for a second-stage regression of efficiency estimates on environmental variables Z to estimate anything sensible and meaningful. If (5.3) does not hold, then such second-stage regressions estimate some unknown thing that is not a feature of any model in which the left-hand side efficiency estimates estimate distance to the frontier of the production set. If (5.3) does not hold, then the FDH, VRS-DEA or CRS-DEA estimators of (2.3)–(2.5) estimate distance to a frontier that may not be attainable and is not relevant.⁵

The separability condition is an important research question whenever environmental variables are available. A formal test of separability versus non-separability is discussed below in Section 5.2. If the condition (5.3) is rejected, the model is said to be “non-separable,” and in this case conditional efficiency estimators are needed to estimate distance to the boundary of Ψ^z . If the model is non-separable, then the unconditional efficiency estimators discussed above in Section 3 estimate distance to the boundary of Ψ instead of Ψ^z , but $\Psi^z \subset \Psi$ for some $z \in \mathcal{Z}$. In addition, conditional efficiency estimators are also needed to build the test of separability below in Section 5.2.

Conditional efficiency measures are defined along the lines of the usual “marginal” scores defined above in (2.3)–(2.5). Here, in the conditional case, the attainable set is restricted to Ψ^z . For example, for the input orientation, we have the conditional efficiency measure

$$\theta_c(x, y \mid z) = \theta(x, y \mid \Psi^z) = \inf\{\theta > 0 \mid (\theta x, y) \in \Psi^z\}. \quad (5.4)$$

Since by (5.2), $\forall z, \Psi^z \subseteq \Psi$, we have $\theta(x, y) \leq \theta_c(x, y \mid z) \leq 1$.

⁵Unfortunately, papers continue to appear in which these points are ignored. Such papers should be disregarded.

Nonparametric estimators are based on an iid sample $\mathcal{S}_n = \{(X_i, Y_i, Z_i)\}_{i=1}^n$ of observations on input and output quantities and corresponding environmental variables. The usual, unconditional FDH and DEA estimators are computed by estimating Ψ , the marginal support of (X, Y) , by the appropriate envelopment of the marginal cloud of points $\mathcal{X}_n = \{(X_i, Y_i)\}_{i=1}^n$ as in (3.1)–(3.3). Here, it is important to make the reference sample used for estimation explicit in the notation. Hence we write, as above for the marginal (input oriented) estimators,

$$\hat{\theta}(X_i, Y_i) = \hat{\theta}(X_i, Y_i \mid \mathcal{X}_n) \quad (5.5)$$

and similarly for the output and hyperbolic orientations.

The conditional efficiency estimators require FDH (or VRS-DEA or CRS-DEA) estimators of Ψ^z . The easiest case is when the variables in Z are discrete, with several firms observed for each value of the discrete variable. In this case, for any possible value z of Z , we have a sample of observations (X, Y) corresponding to the cases where $Z = z$. This defines a subsample $\mathcal{X}_n^z = \{(X_{i_j}, Y_{i_j}) \mid Z_{i_j} = z\}_{j=1}^{n_z} \subset \mathcal{X}_n$ of size $n_z < n$, where $n_z = \sum_{i=1}^n \mathbb{I}(Z_i = z)$ and $\mathbb{I}(\cdot)$ denotes the indicator function. This cloud of “conditional” data points allows estimation of Ψ^z , providing FDH, VRS-DEA or CRS-DEA versions of the conditional efficiency scores by adapting (3.1)–(3.3) to this subsample. Hence, we denote the conditional (input-oriented) estimator of the efficiency of an observation (X_i, Y_i, Z_i) as

$$\hat{\theta}_c(X_i, Y_i \mid Z_i) = \hat{\theta}(X_i, Y_i \mid \mathcal{X}_n^{Z_i}). \quad (5.6)$$

Conditional efficiency estimators have also been developed for the case of continuous Z . In the continuous case, smoothing techniques are needed to “localize” the FDH and DEA estimators so that only observations (X_i, Y_i) having corresponding values of Z_i lying in an appropriate neighborhood of z are used to estimate Ψ^z . For example, we compute the conditional FDH estimator of the unit facing the environmental conditions $Z = z$ by considering only the free disposal hull (or the free disposal and convex hull for DEA) of the data $\mathcal{X}_{n,h}^z = \{(X_i, Y_i) \mid z - h \leq Z_i \leq z + h\}$, where the vector h of “bandwidths” tunes (or controls) the neighborhood of z . The VRS-DEA and CRS-DEA estimators are computed similarly, using only the convex or conical hulls of the free disposal hull of observations (X_i, Y_i) such that the corresponding observations Z_i satisfy $z - h \leq Z_i \leq z + h$. The bandwidths must be small enough to describe units facing environmental conditions similar to z (thereby minimizing some potential bias), but sufficiently large so as to have enough observations in $\mathcal{X}_{n,h}^z$ to avoid excessive variance. As in many nonparametric problems, bandwidths can be chosen to optimize mean integrated square error, which reflects the tradeoff

between bias and variance. As the bandwidths are made smaller (larger), bias is reduced (increased) while variance is increased (reduced). Theoretical results given by Jeong et al. (2010), Daraio et al. (2018), Bădin et al. (2019) and references cited therein establish that the j th optimal bandwidth, corresponding to the j th element of Z , $j = 1, \dots, r$ must be of order $h_j = c_j n^{-1/(r+4)}$. This leads to conditional efficiency estimators with rate of convergence $(nh^r)^\kappa = n^{4\kappa/(r+4)} < n^\kappa$, where $\kappa = 1/(p+q)$, $2/(p+q+1)$ or $2/(p+q)$ for conditional FDH, VRS-DEA or CRS-DEA estimators, respectively. For technical details and asymptotic properties, see Jeong et al. (2010), Daraio et al. (2018) and Bădin et al. (2019). Daraio et al. (2018) provide CLTs for means of conditional efficiency scores that are used to derive the separability test for continuous Z .

5.2 Testing Separability

Simar and Wilson (2007) noted that the separability condition in (5.3) is very restrictive and should be tested, but a development of a formal test had to wait for the additional results of Kneip et al. (2015b). Daraio et al. (2018) developed CLTs for conditional efficiency measures, and used these to develop a test of separability for general cases where Z is either continuous or discrete.⁶ To simplify we explain here the basic ideas of the test in the presence of a single discrete variable Z . To establish necessary notation, let $(z_1, \dots, z_K)$ be the set of possible values for the variable Z . Then the random sample $\mathcal{S}_n$ provides n_k units with $Z_i = z_k$ for $k = 1, \dots, K$ with $n_1 + \dots + n_K = n$. The applied researcher has the marginal cloud of points in $\mathcal{X}_n$ as well as the conditional clouds of points $\mathcal{X}_n^k = \{(X_{i_j}, Y_{i_j}) \mid Z_{i_j} = z_k\}_{j=1}^{n_k}$ for $k = 1, \dots, K$.

The idea is to build a test statistic based on the difference between $\mu_\theta = \mathbb{E}(\theta(X, Y))$, the mean of the efficiency defined relative to the support of the marginal set Ψ , and $\mu_{\theta,c} = \mathbb{E}(\theta_c(X, Y \mid Z))$, the mean of the conditional efficiency defined relative to the support of the conditional sets Ψ^{z_k} , $k = 1, \dots, K$. Under the null hypothesis $H_0: \Psi^{z_k} = \Psi$, $\forall k = 1, \dots, K$, the difference $\mu_\theta - \mu_{\theta,c}$ is zero. In addition, as noted by Daraio et al. (2018), we know that $\mu_{\theta,c} \geq \mu_\theta$, and hence so a test statistics can be derived from the difference $\hat{\mu}_{n,\theta,c} - \hat{\mu}_{n,\theta}$ between the corresponding sample means. These sample means require appropriate estimators of the two population means, where by construction $\hat{\mu}_{n,\theta,c} - \hat{\mu}_{n,\theta} \geq 0$. The null hypothesis of separability should be rejected if this difference is “too big.”

Estimates of individual unconditional efficiencies are given by $\hat{\theta}(X_i Y_i \mid \mathcal{X}_n)$, $i = 1, \dots, n$,

⁶The separability test with discrete environmental variables appears in Daraio et al. (2018, Appendix C).

and for the conditional case we have by (5.6)

$$\widehat{\theta}_c(X_i Y_i \mid Z_i) = \widehat{\theta}(X_i Y_i \mid \mathcal{X}_n^k) \quad (5.7)$$

for $i = 1, \dots, n$ and k such that $Z_i = z_k$. Here either FDH, VRS-DEA or CRS-DEA estimates can be used as appropriate. Similar expressions exist for other orientations (e.g., output or hyperbolic). However, a test statistic constructed from sample means of the two sets of estimates over $i = 1, \dots, n$ cannot be used, since under the null, the asymptotic distribution of $\widehat{\mu}_{n,\theta,c} - \widehat{\mu}_{n,\theta}$ will be degenerate with mass one at zero as shown by Daraio et al. (2018).

The solution to this problem is analogous to the method used for testing the convexity of Ψ described in Section 4.2. We can randomly split the sample $\mathcal{S}_n$ into two parts $\mathcal{S}_{1,n_1}$ and $\mathcal{S}_{2,n_2}$ such that $n_1 = \lfloor n/2 \rfloor$, $n_2 = n - n_1$ such that $\mathcal{S}_{1,n_1} \cup \mathcal{S}_{2,n_2} = \mathcal{S}_n$ and $\mathcal{S}_{1,n_1} \cap \mathcal{S}_{2,n_2} = \emptyset$.⁷ The n_1 observations in $\mathcal{S}_{1,n_1}$ are used to provide the unconditional estimates, while the n_2 observations in $\mathcal{S}_{2,n_2}$ are used to provide the conditional estimators. The sample $\mathcal{S}_{1,n_1}$ provides the marginal cloud of points $\mathcal{X}_{1,n_1}$ and the sample $\mathcal{S}_{2,n_2}$ defines the K conditional clouds of points $\mathcal{X}_{2,n_2}^k = \{(X_{i_j}, Y_{i_j}) \mid Z_{i_j} = z_k\}_{j=1}^{n_{2,k}}$ for $k = 1, \dots, K$, where $(X_{i_j}, Y_{i_j}, Z_{i_j}) \in \mathcal{S}_{2,n_2}$.

Estimators of the unconditional and conditional means are given by

$$\widehat{\mu}_{n_1} = n_1^{-1} \sum_{(X_i, Y_i, Z_i) \in \mathcal{S}_{1,n_1}} \widehat{\theta}(X_i Y_i \mid \mathcal{X}_{1,n_1}) \quad (5.8)$$

and

$$\widehat{\mu}_{c,n_2} = n_2^{-1} \sum_{(X_i, Y_i, Z_i) \in \mathcal{S}_{2,n_2}} \widehat{\theta}(X_i Y_i \mid \mathcal{X}_{2,n_2}^{Z_i}). \quad (5.9)$$

As noted by Daraio et al. (2018), the conditional estimates in group 2 have the usual convergence rate of the unconditional efficiency estimates in group 1. This is because for discrete Z , no bandwidth is involved in the estimation, and each subsample $\mathcal{X}_{2,n_2}^k$ has a size proportional to n_2 , for $k = 1, \dots, K$. Splitting the sample ensures that the two sample means are independent, avoiding difficulties that would otherwise arise from any nonzero covariance.

Given the independent sample means in (5.8) and (5.9), a statistic for testing separability can be constructed along the lines of the difference-in-means test described above in Section 4.1 by adapting the test statistics defined in (4.2) and (4.3). The bias correction terms for group 2 must be computed separately and independently for each subgroup $k = 1, \dots, K$.

⁷Alternatively, one can stratify the splitting, by splitting each subgroups of data $\mathcal{X}_n^k$ independently for $k = 1, \dots, K$.

This involve splitting each of the K subgroups of group 2 to compute the generalized jackknife estimates of bias for each subgroup as described above. See Daraio et al. (2018) for additional details and proofs. Conditional efficiency estimators, as well as the separability tests of Daraio et al. (2018), are implemented in the FEAR software described by Wilson (2008).

6 Dealing with Dynamic Issues

Malmquist productivity indices are widely used in dynamic contexts for measurement of productivity change from one period to another. They are often estimated using nonparametric, data envelopment analysis (DEA) estimators due to the work of Färe et al. (1992). Kneip et al. (2020) provide the theoretical developments needed to make inference about productivity change measured by Malmquist indices when Ψ is convex, for both individual producers as well as for geometric means over a sample of producers. Malmquist indices are often decomposed into various measures of the sources of productivity change, such as changes in efficiency or changes in technology (i.e., movement of the efficient frontier over time). Simar and Wilson (2019) provide CLTs that permit inference about geometric means of the components of Malmquist indices, as well as arithmetic means of their logarithms. We summarize here the results providing asymptotic confidence intervals for changes in efficiency and for changes in technology that are used in our application below in Section 7. For more details, see Simar and Wilson (2019) and Kneip et al. (2020).

Assume that at time t , the attainable set is given by

$$\Psi^t = \{(x, y) \mid x \text{ can produce } y \text{ at time } t\}. \quad (6.1)$$

Then the efficient frontier at time t is given by

$$\Psi^{t\theta} = \{(x, y) \in \Psi^t \mid (\gamma^{-1}x, \gamma y) \notin \Psi^t \forall \gamma > 1\}. \quad (6.2)$$

As in Section 2, we assume the same, usual characteristics for Ψ^t coming from production theory.

Now consider a sample of inputs and outputs for n units observed in periods $t = 1$ and $t = 2$. Denoting $W_i^t = (X_i^t, Y_i^t)$, the sample is represented by $\mathcal{X}_n = \{(W_i^1, W_i^2)\}_{i=1}^n$. Note that the W_i^t are independent for $i = 1, \dots, n$, but may be dependent across time, i.e., W_i^1 and W_i^2 may be dependent. For notational simplicity, assume here and below that the same units are observed in both time periods (Simar and Wilson, 2019 consider the more general situation where some units are observed in only one period). We analyze the components

of the Malmquist indices described in equation (2.11) in Simar and Wilson (2019), in term of hyperbolic measures of efficiency defined in (2.5). Working in the hyperbolic orientation avoids problems that may arise when working in the input or output directions, where some cross-period efficiencies might not be defined. The change in efficiency for the i th unit from t_1 to t_2 is measured by

$$\mathcal{E}_i = \frac{\gamma(W_i^2 | \Psi^2)}{\gamma(W_i^1 | \Psi^1)}, \quad (6.3)$$

and the technology change from t_1 to t_2 faced by the i th unit is described by

$$\mathcal{T}_i = \left(\frac{\gamma(W_i^2 | \Psi^1)}{\gamma(W_i^2 | \Psi^2)} \times \frac{\gamma(W_i^1 | \Psi^1)}{\gamma(W_i^1 | \Psi^2)} \right)^{1/2}. \quad (6.4)$$

Note that both $\mathcal{E}_i$ and $\mathcal{T}_i$ provide measures of change regardless of whether the attainable sets are convex or whether returns to scale are constant or otherwise. Consequently, estimators $\widehat{\mathcal{E}}_i$ and $\widehat{\mathcal{T}}_i$ of $\mathcal{E}_i$ and $\mathcal{T}_i$ can be obtained by replacing the hyperbolic distance measures on the right-hand sides of (6.3) and (6.4) with FDH, VRS-DEA or CRS-DEA estimators, depending what is assumed about Ψ^t for $t \in \{1, 2\}$. Then the convergence rates for the resulting estimators of $\mathcal{E}_i$ and $\mathcal{T}_i$ are the same as the rate for the particular estimator substituted on the right-hand side.

Simar and Wilson (2019) extend the results of Kneip et al. (2020) to derive the asymptotic behavior of these estimators which share the same rate of convergence n^κ of the efficiency estimators used to construct $\widehat{\mathcal{E}}_i$ and $\widehat{\mathcal{T}}_i$, where corresponding values of κ are given in Section 3 and depend on the particular estimator used (i.e., FDH, VRS-DEA or CRS-DEA). These results permit estimation of asymptotic confidence intervals for $\mathcal{E}_i$ and $\mathcal{T}_i$ using the subsampling methods described in Simar and Wilson (2011a).

Researchers often report geometric means of estimates of changes in efficiency or technology over all of the n units represented in a given sample. For efficiency change, applied researchers may compute the geometric mean

$$\widehat{G}_{\mathcal{E},n} = \left[\prod_{i=1}^n \widehat{\mathcal{E}}_i \right]^{1/n} = \exp \left(n^{-1} \sum_{i=1}^n \log(\widehat{\mathcal{E}}_i) \right), \quad (6.5)$$

which can be viewed as an estimator of $G_{\mathcal{E}} = \exp(\mathbb{E}[\log(\mathcal{E}_i)])$. The geometric mean of technical change estimates $\widehat{\mathcal{T}}_i$, $i = 1, \dots, n$ can be computed similarly. Simar and Wilson (2019, Theorem 4.10) derive limit theorems for these geometric means and other variants. These CLTs are analogous, *mutatis mutandis*, to the CLTs presented in (4.2) and (4.3), with limiting normal distributions with the rate of convergence governed by κ (and with the

value of κ varying according the chosen nonparametric estimator of the individual efficiencies and the dimension $p + q$ of the production set).

For example, for the geometric mean of the estimated efficiency changes, we have

$$\sqrt{n} \left(\widehat{G}_{\mathcal{E},n} - G_{\mathcal{E}} - \widehat{B}_{\mathcal{E},n,\kappa} - R_{\mathcal{E},n,\kappa} \right) \xrightarrow{\mathcal{L}} N(0, G_{\mathcal{E}}^2 \sigma_{\mathcal{E}}^2), \text{ with } R_{\mathcal{E},n,\kappa} = o(n^{-1/2}) \quad (6.6)$$

as $n \rightarrow \infty$ provided (i) $p + q \leq 3$ if the FDH estimator is used; (ii) $p + q \leq 4$ and Ψ^t is convex if the VRS-DEA estimator is used; or (iii) $p + q \leq 5$ if Ψ^t is convex, CRS holds, and either the VRS-DEA or CRS-DEA estimator is used. Alternatively, if $\kappa < 1/2$ then

$$\sqrt{n^\kappa} \left(\widehat{G}_{\mathcal{E},n^\kappa} - G_{\mathcal{E}} - \widehat{B}_{\mathcal{E},n,\kappa} - R_{\mathcal{E},n,\kappa} \right) \xrightarrow{\mathcal{L}} N(0, G_{\mathcal{E}}^2 \sigma_{\mathcal{E}}^2), \text{ with } R_{\mathcal{E},n,\kappa} = o(n^{-\kappa}) \quad (6.7)$$

as $n \rightarrow \infty$ under appropriate assumptions on Ψ^t corresponding to the particular estimator used (e.g., FDH, VRS-DEA or CRS-DEA).

The bias correction $\widehat{B}_{\mathcal{E},n,\kappa}$ is computed using a generalized jackknife procedure similar to the one described in Section 3. The sample $\mathcal{X}_n$ must be randomly split into two subsamples as before to ensure independence, but here the split is done across the set of n firms, keeping the same pair of observations for a given firm at times t_1 and t_2 in the same subsample to preserve the time- dependence (see Simar and Wilson, 2019 for details). Since we have consistent estimators of $G_{\mathcal{E}}$ and of $\sigma_{\mathcal{E}}^2$, the limit theorems in (6.6) and (6.7) can be used to construct asymptotic confidence intervals for $G_{\mathcal{E}}$ by using quantiles of the normal distribution. Equivalently, as is standard in statistical practice, one can test the null hypothesis that $G_{\mathcal{E}} = 1$ (i.e, no efficiency changes between the two periods) versus the alternative hypothesis that $G_{\mathcal{E}} \neq 1$, by checking whether the appropriate estimated confidence interval includes the value 1. Estimates of the measures of efficiency and technology change in (6.3) and (6.4), as well the Malmquist index and its other components, can be computed using the FEAR library introduced by Wilson (2008).

7 Efficiency of US Municipal Governments

In this section we show how the tools discussed above can be used to analyze the performance of US municipal governments. We examine performance in two years, 2002 and 2012. Data are taken from the Annual Survey of State and Local Government Finances, the Annual U.S. Building and Permit Survey, the U.S. Census of Governments, the U.S. Bureau of Labor Statistics (BLS) and the Federal Bureau of Investigation (FBI) to define input and output variables to define input and output variables broadly representative of those used

in studies of local governments’ efficiency in the literature. All dollar amounts are measured in terms of thousands of constant, 2008 U.S. dollars. We use the U.S. Consumer Price Index (CPI) reported by the BLS for cities and regions to adjust for price differences across municipalities in the U.S.⁸

Studies of local governments are often limited by data availability, and the story here is no different. Nonetheless, we use total current operating expenditures as our single input (denoted by X) so that $p = 1$. This is consistent with other studies, e.g., Štastná and Gregor (2015). We define $q = 6$ output variables consisting of total population (Y_1), total charges for sewerage and waste management (Y_2), the reciprocal of the total crime rate (Y_3), total land area (in square miles) (Y_4), total building permits (Y_5) and the employment rate (Y_6). These are again representative of what others have specified. The broad collection of outputs reflects the wide variety of goods and services provided by municipal governments in the U.S. Total population reflects the number of people served, and is widely used in studies of local governments (e.g., lo Storto, 2013). The reciprocal of the total crime rate reflects law enforcement services providing public safety. Total land area gives a measure of the physical space over which goods and services are provided. Total building permits provides a measure of administrative services provided, and the employment rate gives a measure of amenities (including jobs, but this is also related to other amenities that attract employers). We use the U.S. Census Bureau’s definition of local government, corresponding to type code 2 in the U.S. Census of Governments’ 14-digit government ID code.

After eliminating observations with missing values for one or more of our variables, we have 730 observations for 2002 and 800 observations for 2012, giving a total of 1,530 observations. As one might expect, there is a good deal of collinearity among the six outputs. We standardize the outputs by dividing each variable by its standard deviation, and then compute the eigenvalues of the (raw) moment matrix for the six standardized outputs as discussed in Section 3.4. The ratio of the largest eigenvalue to the sum of the six eigenvalues is 0.9851, providing clear evidence in view of Wilson (2018) that dimension-reduction is likely to reduce the mean-square error of our efficiency estimates. Consequently, we multiply the $(n \times 6)$ matrix of standardized outputs by the eigenvector corresponding to the largest eigenvalue to construct a single output measure denoted by Y^* as discussed above in Section 3.4. In all of our estimation, we use the single input measure X and the single output measure Y^* , and hence $p = q = 1$. Summary statistics for these variables as well as the original six output measures are provided in Table 1. The summary statistics reveal that several of the

⁸Separate versions of the CPI are available for select cities. Where these are not available, we use regional CPIs to adjust for price differences.

marginal distributions are heavily skewed to the right, reflecting the distribution of city sizes in the U.S. Table 2 gives the number of observations for each year by census region; region 1–4 correspond to the Northeastern, Midwestern, Southern and Western parts of the U.S.⁹

We first estimate efficiency for each year across all regions, and then by region using only observations in a specific year and in region $j \in \{1, 2, 3, 4\}$ to estimate hyperbolic efficiency given by (2.5) for observations in the same year and region. We use first the original input-output specification with X and $Y_1, \dots, Y_6$ and then with the reduced-dimension data consisting of X and Y^* , and we consider the FDH, VRS-DEA and CRS-DEA estimators. We report in Table 3 the number of observations with efficiency estimates equal to one in each case. Using the FDH estimator on the original data, almost half of the observations in each year yield estimates equal to one when data are pooled across regions. Estimating independently for each region, the FDH estimator yields estimates equal to one for well over half of the observations in a region-year. The VRS-DEA estimator yields far fewer estimates equal to one, and the CRS-VRS estimator yields still fewer such estimates. This is to be expected, but it is important to realize that if one used only the VRS-DEA estimator, most of the apparent inefficiency one would find is only a consequence of the VRS-DEA estimator’s imposition of convexity on Ψ . But whether Ψ is convex is an empirical question.

Before testing convexity of Ψ , we first test separability with respect to census regions for each year, and then we test separability with respect to time for each census region. For these tests, we use the reduced-dimension data and input, output and hyperbolic oriented FDH estimators. We use the FDH estimators since they require neither convexity nor constant returns to scale, and thus are the most flexible among the three types of nonparametric estimators. We use 100 sample-splits and 1,000 bootstrap replications as described in Section 5.2. Results are given in Tables 4–5, where “Test #1” indicates tests based on averaging test statistics across the sample splits, and “Test #2” indicates tests based on the distribution of the p – values for each split as discussed in Section 5.2. For the tests of separability with respect to census regions we have six tests, and each yields a p -value of zero as shown in Table 4. For the tests of separability with respect to time by census regions, we have 24 tests, and again all of the 24 p values are equal to zero. Consequently, the tests provide strong evidence against separability with respect to either time or region. The evidence suggests that in a given year, municipalities in different regions face different frontiers, and in a given region, these frontiers are changing over time (this will be revisited below).

The results of the separability tests make clear that convexity should be tested for each

⁹A map showing census regions of the U.S. is available at <https://www2.census.gov/programs-surveys/sahie/reference-maps/2014/us-regdiv.jpg>.

region-year. Table 6 gives the results of these tests. Again, we have two tests, #1 and #2, as mentioned above, and we again use 100 sample-splits and 1,000 bootstrap replications. All together, there are 48 tests, and the results are mixed. For example, for region 4 in 2012, convexity is clearly rejected in 5 out of 6 cases; the only case in which convexity is not rejected is Test #1 using the output orientation. We test in all three directions as a check on robustness, and the results illustrate that conflicting results can arise. The failure to reject in the output orientation may be a consequence of the heavy right-skewness of X and Y^* , but in any case, there is substantial evidence against convexity for this region-year. By contrast, we fail to reject convexity for region 1 in 2002. Since we reject convexity in some, although not all cases, we subsequently use only FDH estimators, as they remain consistent in either case, i.e., regardless of whether Ψ is convex, whereas the VRS-DEA and CRS-DEA estimators do not. Moreover, the results of the separability tests discussed above indicate that efficiency should be estimated independently for each region-year.

In Tables 7–9 we give summary statistics for hyperbolic, input and output oriented FDH efficiency estimates in each region-year. These results reveal a wide range of performances by municipalities in each region-year. Comparison across orientations also reveals some differences, and in particular, the hyperbolic measure suggests greater efficiency on average than either the input or output measures. The hyperbolic measure defined in (2.5) is less sensitive to the curvature of the frontier where an inefficient observation is projected onto the frontier than either the input or the output measures for reasons discussed by Wilson (2011). It is worth noting that our input measure X is total operating expenditure, i.e., cost. Consequently, the input-oriented estimates in Table 8 can be viewed as estimates of cost-efficiency as discussed by Simar and Wilson (2020b). While this is true, it is also worth considering other paths to the frontier of the set of feasible costs and output quantities as given in Tables 7 and 9.

Tables 10–11 report geometric means of estimates of the efficiency-change and technology-change measures defined in (6.3) and (6.4), along with corresponding p -values indicating significant differences from 1 as discussed above in Section 6. Once again, the input, output and hyperbolic directions are used as a robustness check. In Table 10, the results give clear evidence of efficiency change in region 1. For the input and hyperbolic orientations, values greater than 1 indicate an increase in inefficiency, while for the output orientation, values less than 1 indicate an increase in inefficiency. Each of the three estimates for region 1 indicate a significant increase in efficiency from 2002 to 2012. Tests using the output orientation are insignificant for regions 2–4, and for region 3 using the hyperbolic orientation. Results using

the input orientation are significant for each region, and using the hyperbolic orientation leads to significant results not just for region 1 but also for regions 2 and 4. For region 2, the input and hyperbolic directions agree and suggest a decline in efficiency from 2002 to 2012. The results for regions 3 and 4 are mixed, and so one should be careful about drawing conclusions for these regions as the evidence is not clear.

Estimation of the technical change measure defined in (6.4) requires estimation of distances from observations in one period to the frontier in another period as revealed by the expression in (6.4). To avoid numerical problems arising from not being able to reach the frontier from a different period in the input or output directions, we estimate the technical change measure using only the hyperbolic direction. The results in Table 9 reveal statistics that are significantly less than one for each of the four regions, indicating that technology shifted downward from 2002 to 2012. The separability tests in Table 5 indicate that the technology is *different* in 2002 and 2012 for each region, but the results here give evidence on the *direction* of the change. For region 1, the improvement in efficiency suggested by results in Table 10 is offset to some degree by the downward shift of the efficient frontier. While municipalities in region 1 may have increased their efficiency, this was due in part to the downward shift in the technology as opposed to an increase in productivity.

8 Conclusions

The discussion of techniques and methods in this chapter are, for purposes of brevity, limited to radial (i.e., proportional) efficiency measures. However, all of the results discussed above extend to directional efficiency measures developed by Chambers et al. (1996, 1998). With directional measures, efficiency of a unit is measured by its distance to the efficient boundary in a direction fixed by a given vector in $\mathbb{R}_+^{p+q}$; efficiency is given by an additive quantity as opposed to a multiplicative factor as in the case of radial measures, but otherwise the details and associated statistical issues are similar. Asymptotic properties of FDH estimators of these directional distances are developed by Simar and Vanhems (2012), and by Simar et al. (2012) for corresponding DEA estimators. The CLTs developed by Kneip et al. (2015b) and Daraio et al. (2018) as well as the tests developed by Kneip et al. (2016) and Daraio et al. (2018) described above extend easily to additive, directional distances.

It is well-known that envelopment estimators such as those discussed in this chapter are susceptible to outliers in the data. As discussed by Wilson (1993) and others, outliers can be observations from the tail of a distribution, and hence important to consider, or they can result from corrupted data in which case they should be removed. Various methods for

detecting outliers have been developed, including those described by Wilson (1993, 1995), Simar (2003) and Porembski et al. (2005). Finding outliers in more than a few dimensions is a formidable task, and more than one method should be used. Alternatively, one can use robust (with respect to outliers) methods based on order- α frontiers (e.g., see Aragon et al., 2005; Daouia and Simar, 2007; and Wheelock and Wilson, 2008) or order- m frontiers (e.g., see Cazals et al., 2002). Both of these ideas are based on the idea of estimating not the full frontier Ψ^θ , but rather a feature (i.e., a “partial frontier”) lying close to the full frontier. Order- α frontiers are based on quantiles, while order- m frontiers use a trimming parameter (denoted by m) to eliminate the influence of potential outliers or extreme points in the data. Both methods provide benchmarks against which efficiency can be measured. See Simar and Wilson (2013) for additional discussion.

Some research also investigates how noise in the observation process might be accounted for, along the lines of the so-called stochastic frontier analysis, but without introducing restrictive functional-form assumptions. The literature on stochastic frontier analysis involves fully-parametric approaches, but this new strain in the literature focuses on nonparametric and semiparametric models and methods. So far, most of the work in this area is limited to univariate response models (e.g., production functions or cost functions); e.g., see the work by Kumbhakar et al. (2007), Kneip et al. (2015a), Simar et al. (2017) and Florens et al. (2020). Multivariate extensions include Simar (2007), Simar and Zelenyuk (2011) and Simar and Wilson (2021).

For too many years, FDH and DEA estimators were used without any statistical underpinnings. Papers in the literature from the 1980s and beyond often imply that truth can be learned from data, but in fact, one can only obtain estimates and make inference using data. Today, however, empirical researchers can bring a “tool box” full of statistical tools to data in order to conduct rigorous analyses and to quantify the uncertainty surrounding estimates can be quantified. The various testing methods discussed here have been implemented in the FEAR software package introduced by Wilson (2008), and so there is little excuse for not using these methods.

References

- Adler, N. and B. Golany (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.
- Afonso, F. S. (2008), Assessing and explaining the relative efficiency of local government, *Journal of Socio-Economics* 37, 1946–1979.
- Afriat, S. (1972), Efficiency estimation of production functions, *International Economic Review* 13, 568–598.
- Aigner, D., C. A. K. Lovell, and P. Schmidt (1977), Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics* 6, 21–37.
- Aragon, Y., A. Daouia, and C. Thomas-Agnan (2005), Nonparametric frontier estimation: A conditional quantile-based approach, *Econometric Theory* 21, 358–389.
- Bădin, L., C. Daraio, and L. Simar (2012), How to measure the impact of environmental factors in a nonparametric production model, *European Journal of Operational Research* 223, 818–833.
- (2014), Explaining inefficiency in nonparametric production models: The state of the art, *Annals of Operations Research* 214, 5–30.
- (2019), A bootstrap approach for bandwidth selection in estimating conditional efficiency measures, *European Journal of Operational Research* 277, 784–797.
- Banker, R. D., A. Charnes, and W. W. Cooper (1984), Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science* 30, 1078–1092.
- 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 (1996), Benefit and distance functions, *Journal of Economic Theory* 70, 407–419.
- (1998), Profit, directional distance functions, and nerlovian efficiency, *Journal of Optimization Theory and Applications* 98, 351–364.
- Charnes, A., W. W. Cooper, and E. Rhodes (1978), Measuring the efficiency of decision making units, *European Journal of Operational Research* 2, 429–444.
- Da Cruz, N. F. and R. C. Marques (2014), Revisiting the determinants of local government performance, *Omega* 44, 91–103.
- 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.

- (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis*, New York: Springer Science+Business Media, LLC.
- Daraio, C., L. Simar, and P. W. Wilson (2018), Central limit theorems for conditional efficiency measures and tests of the ‘separability condition’ in non-parametric, two-stage models of production, *The Econometrics Journal* 21, 170–191.
- De Borger, B. and K. Kerstens (2000), What is known about municipal efficiency? the Belgian case and beyond, in J. Blank, ed., *Public Provision and Performance: Contributions from Efficiency and Productivity Measurement*, Amsterdam: Elsevier, pp. 299–330.
- de Oliveira Junqueira, M. (2015), Efficiency measurement of local public sector in international perspective: A literature survey. Unpublished working paper, University of São Paulo, Brazil. Presented at International Conference on Public Policy, Milan, 2015.
- 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.
- Ellickson, B. (1971), Jurisdictional fragmentation and residential choice, *American Economic Review* 61, 334–339.
- Färe, R., S. Grosskopf, B. Lindgren, and P. Roos (1992), Productivity changes in Swedish pharmacies 1980–1989: A non-parametric Malmquist approach, *Journal of Productivity Analysis* 3, 85–101.
- Färe, R., S. Grosskopf, and C. A. K. Lovell (1985), *The Measurement of Efficiency of Production*, Boston: Kluwer-Nijhoff Publishing.
- Farrell, M. J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society A* 120, 253–281.
- Florens, J. P., L. Simar, and I. Van Keilegom (2020), Estimation of the boundary of a variable observed with symmetric error, *Journal of the American Statistical Association* 115, 425–441.
- Grosskopf, S., K. J. Hayes, L. L. Taylor, and W. L. Weber (2001), On the determinants of school district efficiency: Competition and monitoring, *Journal of Urban Economics* 49, 453–478.
- Grossman, P. J., P. Mavros, and R. W. Wassmer (1999), Public sector technical inefficiency in large U.S. cities, *Journal of Urban Economics* 46, 278–299.
- Härdle, W. and L. Simar (2019), *Applied Multivariate Statistical Analysis*, Berlin: Springer-Verlag. Fifth edition.
- Hayes, K. J., L. Razzolini, and L. B. Ross (1998), Bureaucratic choice and nonoptimal provision of public goods: Theory and evidence, *Public Choice* 94, 1–20.
- Henderson, J. M. and R. E. Quandt (1971), *Microeconomic Theory: A Mathematical Approach*, New York: McGraw-Hill Book Company.

- 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.
- (2020), Inference in dynamic, nonparametric models of production: Central limit theorems for Malmquist indices, *Econometric Theory* Forthcoming.
- 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 (2007), Nonparametric stochastic frontiers: A local likelihood approach, *Journal of Econometrics* 137, 1–27.
- Lesourne, J. (1972), *Le Calcul Économique*, Paris: Dunod.
- lo Storto, C. (2013), Evaluating technical efficiency of Italian major municipalities: a data envelopment analysis model, *Procedia-Social and Behavioral Sciences* 81, 346–350.
- Mas-Colell, A., M. D. Whiston, and J. R. Green (1995), *Microeconomic Theory*, New York: Oxford University Press, Inc.
- Meeusen, W. and J. van den Broeck (1977), Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review* 18, 435–444.
- Narbón-Perpiñá, I. and K. De Witte (2018), Local governments’ efficiency: a systematic literature review—part I, *International Transactions in Operational Research* 25, 431–468.
- 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.

- Porembski, M., K. Breitenstein, and P. Alpar (2005), Visualizing efficiency and reference relations in data envelopment analysis with an application to the branches of a German bank, *Journal of Productivity Analysis* 23, 203–221.
- Robinson, P. M. (1988), Root- n -consistent semiparametric regression, *Econometrica* 56, 931–954.
- Simar, L. (2003), Detecting outliers in frontier models: A simple approach, *Journal of Productivity Analysis* 20, 391–424.
- (2007), How to improve the performances of DEA/FDH estimators in the presence of noise, *Journal of Productivity Analysis* 28, 183–201.
- Simar, L., I. Van Keilegom, and V. Zelenyuk (2017), Nonparametric least squares methods for stochastic frontier models, *Journal of Productivity Analysis* 47, 189–204.
- Simar, L. and A. Vanhems (2012), Probabilistic characterization of directional distances and their robust versions, *Journal of Econometrics* 166, 342–354.
- Simar, L., A. Vanhems, and P. W. Wilson (2012), Statistical inference for DEA estimators of directional distances, *European Journal of Operational Research* 220, 853–864.
- Simar, L. and P. W. Wilson (2007), Estimation and inference in two-stage, semi-parametric models of productive efficiency, *Journal of Econometrics* 136, 31–64.
- (2011a), Inference by the m out of n bootstrap in nonparametric frontier models, *Journal of Productivity Analysis* 36, 33–53.
- (2011b), Two-Stage DEA: Caveat emptor, *Journal of Productivity Analysis* 36, 205–218.
- (2013), Estimation and inference in nonparametric frontier models: Recent developments and perspectives, *Foundations and Trends in Econometrics* 5, 183–337.
- (2015), Statistical approaches for non-parametric frontier models: A guided tour, *International Statistical Review* 83, 77–110.
- (2019), Central limit theorems and inference for sources of productivity change measured by nonparametric Malmquist indices, *European Journal of Operational Research* 277, 756–769.
- (2020a), Hypothesis testing in nonparametric models of production using multiple sample splits, *Journal of Productivity Analysis* 53, 287–303.
- (2020b), Technical, allocative and overall efficiency: Estimation and inference, *European Journal of Operational Research* 282, 1164–1176.
- (2021), Nonparametric, stochastic frontier models with multiple inputs and outputs. Unpublished working paper, Institut de Statistique Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
- Simar, L. and V. Zelenyuk (2011), Stochastic FDH/DEA estimators for frontier analysis, *Journal of Productivity Analysis* 36, 1–20.
- Štastná, L. and M. Gregor (2015), Public sector efficiency in transition and beyond: evidence from czech local governments, *Applied Economics* 47, 680–699.

- Tang, K.-L. (1997), Efficiency in the private sector: A critical review of empirical evidence from public services, *International Review of Administrative Sciences* 63, 459–474.
- Tiebout, C. M. (1956), A pure theory of local public expenditure, *Journal of Political Economy* 64, 416–424.
- Tran, C.-D. T. T. and B. Dollery (2020), An empirical analysis of municipal efficiency and local resident satisfaction by council type in Victorian local government, *Australian Journal of Public Administration* 79, 531–549.
- Varian, H. R. (1978), *Microeconomic Analysis*, New York: W. W. Norton & Co.
- 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.

Table 1: Summary Statistics

Variable	Min	Q1	Median	Mean	Q3	Max
X	9.4800(2)	5.7970(4)	1.0492(5)	2.6124(5)	2.2749(5)	1.0844(7)
Y_1	7.9700(2)	2.9383(4)	4.5486(4)	8.6136(4)	7.7029(4)	3.7926(6)
Y_2	1.0381(1)	7.6186(3)	1.4982(4)	3.1267(4)	2.8571(4)	1.0198(6)
Y_3	2.8500(-6)	1.7451(-4)	2.5331(-4)	3.2229(-4)	3.7059(-4)	9.5493(-3)
Y_4	5.4000(-1)	1.1950(1)	2.1440(1)	4.2049(1)	3.8172(1)	2.7167(3)
Y_5	1.0000(0)	4.9000(1)	1.7300(2)	4.1293(2)	4.4800(2)	8.8460(3)
Y_6	8.0800(1)	9.1900(1)	9.3600(1)	9.3235(1)	9.5000(1)	9.8500(1)
Y^*	8.1739(2)	3.0965(4)	4.8038(4)	9.1560(4)	8.0390(4)	3.9234(6)

NOTE: Values are given in scientific notation; e.g., $8.1739(2)$ represents $8.1739 \times 10^2 = 817.39$.

Table 2: Sample Sizes by Year and U.S. Census Region

Year	All	Region 1	Region 2	Region 3	Region 4
	Regions				
2002	730	75	158	289	208
2012	800	72	200	299	229
TOTAL	1530	147	358	588	437

Table 3: Numbers of Observations With Estimated Hyperbolic Technical Efficiency Equal to 1 in Each Year

Year	Region	n	Without Dimension — Reduction —			With Dimension — Reduction —		
			FDH	VRS	CRS	FDH	VRS	CRS
2002	all	730	364	35	6	28	5	1
2012	all	800	381	35	7	31	6	1
2002	1	75	61	17	5	9	3	1
	2	158	99	23	4	23	6	1
	3	289	186	44	6	27	7	1
	4	208	148	27	8	19	5	1
2012	1	72	64	23	6	10	3	1
	2	200	118	27	4	23	9	1
	3	299	186	42	5	26	5	1
	4	229	167	33	13	25	5	1

Table 4: Test for Separability with Respect to Census Regions, by Time (100 Splits, 1,000 Bootstrap Replications; $p = q = 1$)

Year	— Input —		— Output —		— Hyperbolic —	
	Statistic	p -value	Statistic	p -value	Statistic	p -value
Test #1:						
2002	9.301	0.000	7.157	0.000	7.806	0.000
2012	11.021	0.000	9.131	0.000	11.814	0.000
Test #2:						
2002	0.919	0.000	0.947	0.000	0.928	0.000
2012	0.997	0.000	0.987	0.000	0.968	0.000

Table 5: Test for Separability with Respect to Time by Census Regions (100 Splits, 1,000 Bootstrap Replications; $p = q = 1$)

Region	— Input —		— Output —		— Hyperbolic —	
	Statistic	p -value	Statistic	p -value	Statistic	p -value
Test #1:						
1	0.182	0.000	5.389	0.000	3.842	0.000
2	0.805	0.000	2.466	0.000	1.636	0.000
3	1.035	0.000	6.015	0.000	6.779	0.000
4	9.527	0.000	15.129	0.000	22.608	0.000
Test #2:						
1	0.333	0.000	0.731	0.000	0.608	0.000
2	0.382	0.000	0.564	0.000	0.478	0.000
3	0.443	0.000	1.000	0.000	1.000	0.000
4	1.000	0.000	0.986	0.000	0.984	0.000

Table 6: Test for Convexity by Year and Census Region (100 Splits, 1,000 Bootstrap Replications; $p = q = 1$)

Year	Region	— Input —		— Output —		— Hyperbolic —	
		Statistic	p -value	Statistic	p -value	Statistic	p -value
Test #1:							
2002	1	0.415	0.768	1.307	0.995	0.212	0.220
	2	0.712	0.752	0.683	0.306	1.234	0.931
	3	0.012	0.555	0.040	0.666	0.171	0.544
	4	1.896	0.000	0.678	0.966	1.801	0.000
2012	1	0.903	0.916	0.022	0.796	1.549	0.998
	2	0.637	0.209	1.139	0.030	0.513	0.772
	3	1.067	0.001	0.274	0.577	0.242	0.041
	4	1.177	0.000	1.966	1.000	1.961	0.000
Test #2:							
2002	1	0.273	0.156	0.374	0.109	0.233	0.347
	2	0.318	0.257	0.340	0.293	0.383	0.141
	3	0.163	0.695	0.187	0.534	0.139	0.829
	4	0.559	0.001	0.318	0.076	0.537	0.000
2012	1	0.359	0.063	0.266	0.445	0.457	0.003
	2	0.513	0.009	0.477	0.019	0.326	0.247
	3	0.504	0.001	0.228	0.509	0.245	0.504
	4	0.552	0.000	0.594	0.000	0.707	0.000

Table 7: Summary Statistics for Hyperbolic FDH Technical Efficiency Estimates by Year and Census Region (with Dimension Reduction; $p = q = 1$)

Year	Min	Q1	Median	Mean	Q3	Max	Var
— Region 1: Northeast —							
2002	0.454	0.566	0.669	0.705	0.884	1.000	0.033
2012	0.407	0.650	0.741	0.764	0.916	1.000	0.028
— Region 2: Midwest —							
2002	0.488	0.729	0.825	0.825	0.949	1.000	0.018
2012	0.370	0.722	0.800	0.797	0.891	1.000	0.018
— Region 3: South —							
2002	0.272	0.604	0.742	0.728	0.860	1.000	0.031
2012	0.275	0.617	0.767	0.747	0.890	1.000	0.029
— Region 4: West —							
2002	0.278	0.561	0.677	0.694	0.837	1.000	0.031
2012	0.301	0.539	0.696	0.701	0.855	1.000	0.038

Table 8: Summary Statistics for Input-Oriented FDH Technical Efficiency Estimates by Year and Census Region (with Dimension Reduction; $p = q = 1$)

Year	Min	Q1	Median	Mean	Q3	Max	Var
— Region 1: Northeast —							
2002	0.190	0.297	0.415	0.525	0.739	1.000	0.072
2012	0.195	0.369	0.590	0.611	0.829	1.000	0.068
— Region 2: Midwest —							
2002	0.166	0.477	0.640	0.667	0.897	1.000	0.053
2012	0.104	0.440	0.584	0.609	0.773	1.000	0.052
— Region 3: South —							
2002	0.046	0.302	0.513	0.530	0.737	1.000	0.073
2012	0.049	0.311	0.515	0.545	0.761	1.000	0.072
— Region 4: West —							
2002	0.028	0.201	0.349	0.439	0.637	1.000	0.081
2012	0.039	0.242	0.361	0.472	0.731	1.000	0.086

Table 9: Summary Statistics for Output-Oriented FDH Technical Efficiency Estimates by Year and Census Region (with Dimension Reduction; $p = q = 1$)

Year	Min	Q1	Median	Mean	Q3	Max	Var
— Region 1: Northeast —							
2002	0.209	0.368	0.481	0.559	0.750	1.000	0.063
2012	0.153	0.352	0.571	0.604	0.876	1.000	0.080
— Region 2: Midwest —							
2002	0.248	0.571	0.725	0.718	0.893	1.000	0.040
2012	0.174	0.557	0.688	0.688	0.814	1.000	0.039
— Region 3: South —							
2002	0.105	0.410	0.593	0.597	0.785	1.000	0.058
2012	0.127	0.424	0.623	0.619	0.815	1.000	0.057
— Region 4: West —							
2002	0.101	0.435	0.547	0.593	0.735	1.000	0.050
2012	0.112	0.402	0.565	0.598	0.776	1.000	0.055

NOTE: Reported values reflect inverse efficiency estimates in order to facilitate comparison with hyperbolic and input-oriented efficiency estimates.

Table 10: Estimates of Efficiency Change, 2002–2012, by Region (FDH; $p = q = 1$)

Region	— Input —		— Output —		— Hyperbolic —	
	Statistic	p -value	Statistic	p -value	Statistic	p -value
1	1.121	9.668(−7)	0.991	6.389(−2)	1.065	6.564(−12)
2	0.968	5.700(−6)	1.009	1.954(−1)	0.991	4.226(−2)
3	0.984	7.298(−3)	0.997	6.075(−1)	1.001	3.184(−1)
4	1.048	2.953(−5)	1.036	9.609(−1)	0.970	2.480(−3)

NOTE: Values in parentheses give value of exponent in scientific notation; e.g., 9.668×10^{-7} is denoted by 9.668(−7).

Table 11: Estimates of Technology Change, 2002–2012, by Region (Hyperbolic FDH; $p = q = 1$)

Region	Statistic	p -value
1	0.820	$3.875(-74)$
2	0.916	$1.687(-8)$
3	0.903	$1.914(-91)$
4	0.967	$9.199(-1)$

NOTE: Values in parentheses give value of exponent in scientific notation; e.g., 3.875×10^{-74} is denoted by $3.875(-74)$.